



```
QY 263 GAAPERPTVVGSGSWAHGRTGSDRGFCVVSAPARAEATSLGALSTGRSHSPVG 322
Db 301 ROHHAGPSTSRPPRPMDTCCPPVYATKHFLLSSGDKEDLRPSFLSSRLPGARL 360
QY 323 ROHHAGPSTSRPPRPMDTCCPPVYATKHFLLSSGDKEDLRPSFLSSRLPGARL 382
Db 361 VERIFLGSRRMPGTPRRRLPQRYQWMPRLFELIGNHACPYGLVLTCHCPLRAVT 420
QY 383 VERIFLGSRRMPGTPRRRLPQRYQWMPRLFELIGNHACPYGLVLTCHCPLRAVT 442
Db 421 PAAGVCAKREKPOGSVAPEEDTDPRRLVOLLRQSSPMOYGFVACLRRLVPPGLMS 480
QY 443 PAAGVCAKREKPOGSVAPEEDTDPRRLVOLLRQSSPMOYGFVACLRRLVPPGLMS 502
Db 481 RHNERRLRNTKFKFISLGNKAKLSLQELTWKMSYRDCAMLRSPGVCVPAAEHRLREEI 540
QY 503 RHNERRLRNTKFKFISLGNKAKLSLQELTWKMSYRDCAMLRSPGVCVPAAEHRLREEI 562
Db 541 LAFLHMLMSVYVELLSRFFYVETTFOKNRLFEYRKSWSKLSIGIRQHLKRVOLRE 600
QY 563 LAFLHMLMSVYVELLSRFFYVETTFOKNRLFEYRKSWSKLSIGIRQHLKRVOLRE 622
Db 601 LSEAEVQRHREARPALTLRLRFPKPDGLRPIVNDYVVGARTFRERKRAERLTSRYKA 660
QY 623 LSEAEVQRHREARPALTLRLRFPKPDGLRPIVNDYVVGARTFRERKRAERLTSRYKA 682
Db 661 LFSVLYNERARRRGLLGASVGLDDIHRAMRTFVLRYRADDPPELTFVYVDTGAYDTI 720
QY 683 LFSVLYNERARRRGLLGASVGLDDIHRAMRTFVLRYRADDPPELTFVYVDTGAYDTI 742
Db 721 PODRLTEVIASITIKPONTYCVRRYAVVOKAAGHVRAKFSHVSTLTDLPYMRQFAHL 780
QY 743 PODRLTEVIASITIKPONTYCVRRYAVVOKAAGHVRAKFSHVSTLTDLPYMRQFAHL 802
Db 781 QETSPLDAVAVIEOSSSLNEASSGLDFVFLRFMCHAAVRIRGKSYVOCGIPGSIILSTL 840
QY 803 QETSPLDAVAVIEOSSSLNEASSGLDFVFLRFMCHAAVRIRGKSYVOCGIPGSIILSTL 862
Db 841 LCSLCYGDMMENKLFAGIRBGLLLRLVDDFLVPHLTHAKTFLRTLVGRVPEYGCYVNL 900
QY 863 LCSLCYGDMMENKLFAGIRBGLLLRLVDDFLVPHLTHAKTFLRTLVGRVPEYGCYVNL 922
Db 901 RKTIVNFPVEDEALGTAFAVQMPAHGL 927
QY 923 RKTIVNFPVEDEALGTAFAVQMPAHGL 949

RESULT 2 PRELIMINARY; PRT; 1132 AA.
ID 014783;
AC 014783;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DE 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE TELOMERASE CATALYTIC SUBUNIT.
OS HEST2.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97433088.
RA MEYERSON M., COUNTER C.M., EATON E.N., ELLISEN L.W., STEINER P.,
RA CADDLE S.D., ZIAUGRA L., BEIJERSBERGEN R.L., DAVIDOFF M.J., LIU Q.,
RA BACCHETTI S., HABER D.A., WEINBERG R.A.;
RL CELL 90:785-795(1997).
DR EMBL: AF018167: G2347129;
SQ SEQUENCE 1132 AA: 126937 MM: C1E5E2AF CRC32;

Query Match 97.7%; Score 6948; DB 4; Length 1132;
Best local similarity 99.9%; Pred. No. 0.00e+00;
Matches 926; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 23 MPAPRCRAVRSLLSRSHYREVLPATFVRLRIGPQGMRLVQRDPAARLVAQCLVCPW 82
Db 61 DARPAPASFRVOSCKELVAVIOLRCEKAGKNVLAFLALDAGRGPPAFTSVR 120
QY 83 DARPAPASFRVOSCKELVAVIOLRCEKAGKNVLAFLALDAGRGPPAFTSVR 142
Db 121 STLPNTVTALRGSGAMGILLRVRGDDVLVHLARCALVVAASCAVOGCPRLYOLGA 180
QY 143 STLPNTVTALRGSGAMGILLRVRGDDVLVHLARCALVVAASCAVOGCPRLYOLGA 202
Db 181 ATQARPPHASGPRRLRGCERAMNHSVREAGVPLGIPAPGARRGGSASRSLPLPRPRR 240
QY 203 ATQARPPHASGPRRLRGCERAMNHSVREAGVPLGIPAPGARRGGSASRSLPLPRPRR 262
Db 241 GAAPERPTVVGSGSWAHGRTGSDRGFCVVSAPARAEATSLGALSTGRSHSPVG 300
QY 263 GAAPERPTVVGSGSWAHGRTGSDRGFCVVSAPARAEATSLGALSTGRSHSPVG 322
Db 301 ROHHAGPSTSRPPRPMDTCCPPVYATKHFLLSSGDKEDLRPSFLSSRLPGARL 360
QY 323 ROHHAGPSTSRPPRPMDTCCPPVYATKHFLLSSGDKEDLRPSFLSSRLPGARL 382
Db 361 VERIFLGSRRMPGTPRRRLPQRYQWMPRLFELIGNHACPYGLVLTCHCPLRAVT 420
QY 383 VERIFLGSRRMPGTPRRRLPQRYQWMPRLFELIGNHACPYGLVLTCHCPLRAVT 442
Db 421 PAAGVCAKREKPOGSVAPEEDTDPRRLVOLLRQSSPMOYGFVACLRRLVPPGLMS 480
QY 443 PAAGVCAKREKPOGSVAPEEDTDPRRLVOLLRQSSPMOYGFVACLRRLVPPGLMS 502
Db 481 RHNERRLRNTKFKFISLGNKAKLSLQELTWKMSYRDCAMLRSPGVCVPAAEHRLREEI 540
QY 503 RHNERRLRNTKFKFISLGNKAKLSLQELTWKMSYRDCAMLRSPGVCVPAAEHRLREEI 562
Db 541 LAFLHMLMSVYVELLSRFFYVETTFOKNRLFEYRKSWSKLSIGIRQHLKRVOLRE 600
QY 563 LAFLHMLMSVYVELLSRFFYVETTFOKNRLFEYRKSWSKLSIGIRQHLKRVOLRE 622
Db 601 LSEAEVQRHREARPALTLRLRFPKPDGLRPIVNDYVVGARTFRERKRAERLTSRYKA 660
QY 623 LSEAEVQRHREARPALTLRLRFPKPDGLRPIVNDYVVGARTFRERKRAERLTSRYKA 682
Db 661 LFSVLYNERARRRGLLGASVGLDDIHRAMRTFVLRYRADDPPELTFVYVDTGAYDTI 720
QY 683 LFSVLYNERARRRGLLGASVGLDDIHRAMRTFVLRYRADDPPELTFVYVDTGAYDTI 742
Db 721 PODRLTEVIASITIKPONTYCVRRYAVVOKAAGHVRAKFSHVSTLTDLPYMRQFAHL 780
QY 743 PODRLTEVIASITIKPONTYCVRRYAVVOKAAGHVRAKFSHVSTLTDLPYMRQFAHL 802
Db 781 QETSPLDAVAVIEOSSSLNEASSGLDFVFLRFMCHAAVRIRGKSYVOCGIPGSIILSTL 840
QY 803 QETSPLDAVAVIEOSSSLNEASSGLDFVFLRFMCHAAVRIRGKSYVOCGIPGSIILSTL 862
Db 841 LCSLCYGDMMENKLFAGIRBGLLLRLVDDFLVPHLTHAKTFLRTLVGRVPEYGCYVNL 900
QY 863 LCSLCYGDMMENKLFAGIRBGLLLRLVDDFLVPHLTHAKTFLRTLVGRVPEYGCYVNL 922
Db 901 RKTIVNFPVEDEALGTAFAVQMPAHGL 927
QY 923 RKTIVNFPVEDEALGTAFAVQMPAHGL 949

RESULT 3 PRELIMINARY; PRT; 1122 AA.
ID 070372;
AC 070372;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DE 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE TELOMERASE REVERSE TRANSCRIPTASE.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
```

OC EUTHERIA; RODENTIA.  
RN (1)  
RP SEQUENCE FROM N.A.  
RA GREENBERG R.A., ALTSOPP R.C., CHIN L., MORIN G.B., DEPINHO R.A.;  
RL ONCOGENE 0:0-0(1998).  
DR EMBL; AF051911; G3005592; -.  
KW RNA-DIRECTED DNA POLYMERASE.  
SQ SEQUENCE 1122 AA; 127977 MW; 222075D6 CRC32;  
  
Query Match 57.58; Score 4087; DB 11; Length 1122;  
Best Local Similarity 61.08; Pred. No. 0.00e+00;  
Matches 574; Conservative 169; Mismatches 163; Indels 35; Gaps 22;  
  
Db 1 MTRAPCAPVARSILRSRYEVLPLATFVRRLGEGRRRLVPGSPKRYRLTYAOCICMH 60  
OY 23 MPAPPCRAVRSILRSRYEVLPLATFVRRLGEGRRRLVPGSPKRYRLTYAOCICMH 82  
Db 61 GSOPPPADLSFHQVSSIKELVAVVQRLCEBNERNYLAFGELLNBARGPPMAFTSSVR 120  
OY 83 DAPPPAPAPSFQVSCLEKELVAVVQRLCEBNERNYLAFGELLNBARGPPMAFTSSVR 142  
Db 121 SYLPNTVIELRYSAGMMLLSRYGDDLYLLAHGALTYLVPPSCAYOVGSPYQICA 180  
OY 143 SYLPNTVIELRYSAGMMLLSRYGDDLYLLAHGALTYLVPPSCAYOVGSPYQICA 202  
Db 181 TTDIWPSVSAVSRPTPVGNFTNLFLQIKSSRQEAAPKPLALSRGKRLSLSTS 240  
OY 203 ATDAPRPPHNS-GRRRLGCE-RA-W--NH--SVR-ENGVLGLPAPGARRGGSSASRS 253  
Db 241 VPSAKKARCPYPRVEGPRQVLPYPSGKSNVSPARS-PEV-PT-AEKDLSR-SGKYS 296  
OY 254 LPLPKPRGAAPERPERTPYGSGMAHPGRTGRPS-DRGCVSPAPAPAEALSLGALS 312  
Db 297 DLSLS-GSYCCCKAPSTSLSPRONAOLRP-FIETRHLYSRGDCERLNPFLSN 354  
OY 313 GTRHSHPSVQROHAGAPRPSRPPWDPRCPRVYAEKHFELYSQD-KQOLPSEFLSS 371  
Db 355 LQNLGARLYEITFLGSRPTSGPLCRTHRLSBRYWOMRPLQOLLVNHAQVRL 414  
OY 372 LKRLSLGARLYEITFLGSRPTSGPLCRTHRLSBRYWOMRPLQOLLVNHAQVRL 431  
Db 415 RSHCRPR--T--AN--Q--O--VT--DALNTSPHMLDLRLHSSPQVGYFRAC 459  
OY 432 KTRCPRLAAVTPAAGVCAKREKPGSVAAPBEEDTDRRLVQLRQSSSPQVGYFRAC 491  
Db 460 CKVVASLWGTNRNERRERFENLKEFISLKGKYLSTQELMKKXVEDCHMLRSSPKDRY 519  
OY 492 RRLVPPGLMGRNERRERFENLKEFISLKGKYLSTQELMKKXVEDCHMLRSSPKDRY 551  
Db 520 PAEHLRLREILATFLFWMMDTYVQQLRSEFYITESTPOKNRLFYRKSVNKLQSIG 579  
OY 552 PAEHLRLREILATFLFWMMDTYVQQLRSEFYITESTPOKNRLFYRKSVNKLQSIG 611  
Db 580 RQLEERLRELSEOEVRHNDTWLAMPICRLFIKPRNGLRPIVNNYSMGSRALGRK 639  
OY 612 RQLEERLRELSEOEVRHNDTWLAMPICRLFIKPRNGLRPIVNNYSMGSRALGRK 671  
Db 640 QAOHFTORLKTLESMYTERTKPHLMGSSVGLMDIYRTWRAVYLRVRLDOTPRMYEV 699  
OY 672 RABRLRSVYALFSLVNYERARRPGLIGASVGLDIDHRAWTFVLRVRODPPLEFV 721  
Db 700 KADVTAGAYDAIPQGLVYVANNIRSESYCTROYAVVARDQOGVHNKFRQVYTTLS 759  
OY 732 KADVTAGAYDAIPQGLVYVANNIRSESYCTROYAVVARDQOGVHNKFRQVYTTLS 790  
Db 760 LQYMGQFHLKLDSDASALRNSVIEOSTISMENSSSLDEFLHLRHSVNVIGRCYV 819  
OY 791 LQYMGQFHLKLDSDASALRNSVIEOSTISMENSSSLDEFLHLRHSVNVIGRCYV 848  
Db 820 QCGGIPGSSLSLTLCSLCEGDMENKLFAYVQDGLLRVDFDLVTPHLDQAKFTLS 879  
OY 849 QCGGIPGSSLSLTLCSLCEGDMENKLFAYVQDGLLRVDFDLVTPHLDQAKFTLS 908

Db 880 LVHGVPEYGCMINLQKTVVNFVEPGETLGAAPYOLPAHCL 920  
OY 909 LVHGVPEYGCYVNLKRTIVNFPVEDALGTAFAVQMPAHGL 949  
  
RESULT 4  
ID 013338 PRELIMINARY; PRT; 988 AA.  
AC 013338;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE TELOMERASE REVERSE TRANSCRIPTASE 1.  
GN TRT1.  
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-972H-;  
RX MEDLINE; 97400623.  
RA NAKAMURA T.M., MORIN G.B., CHAPMAN K.B., WEINRICH S.L., ANDREWS W.H.,  
LINGNER J., HARLEY C.B., CECCH T.R.;  
RL SCIENCE 277:955-959(1997).  
DR EMBL; AF015783; G2340169; -.  
KW RNA-DIRECTED DNA POLYMERASE.  
SQ SEQUENCE 989 AA; 116456 MW; FDE74202 CRC32;  
  
Query Match 9.58; Score 675; DB 3; Length 989;  
Best Local Similarity 26.7%; Pred. No. 2.26e-108;  
Matches 125; Conservative 128; Mismatches 196; Indels 19; Gaps 16;  
  
Db 340 PNOVFALSLIIVRPKILMGNORFELIILMDLEFLKLSYESPSLHLSNLIKISEI 399  
OY 480 PNOVFALSLIIVRPKILMGNORFELIILMDLEFLKLSYESPSLHLSNLIKISEI 539  
Db 400 EMLVLGRKSNAMKSLDEFEKR-KQ-IFAEIYMLYNSFIPILOSFYITESSDLNRTV 457  
OY 540 AML--RRPQGVCAVPAHRLREILAKFLHMLSVYVELLRSEFYITETFOKNRLF 596  
Db 458 YERKDIW-KLCPRTSKMEAEFEKINNNVMDTQ-KTLPRAVIRLLPKKNTFRILT 515  
OY 597 FYKSVWSKLQSIGIRQHLKRVQLRELSEAVRQREARPALTSRLRIPPDGLRPIV 656  
Db 516 NLKRLFLQMSNK-KMLVSTNQLRPAVASTIKHLINDESSGIPRN-LEVYMKLLTFK 573  
OY 657 NM-D-YVVGARFRRKRRARLTSRYKALFVNLNERARRPGLIGASVGLDIDHRAWRT 714  
Db 574 DLLKHMFGRR-K-YEVRIDIKSCYDIRIKODLMFRYKKKLDP-EVIRKYATIH-ATS 629  
OY 715 FYLRVAAQPPPELYVKKDYTGAVDTIPQDLRLTEYIASIIRPQNTYCVRRAYVQKAAH 774  
Db 630 DRATKNVSEAFSYEDMVEER--VYQLLSMT-SDTLFVDFVDYWTSSSEIFKKLKH 686  
OY 775 GHVRKAFKSHVSTFLDQYMGQFVHLOETSPRLRAVYIEQSSSLNEASSGLFVFLRF 834  
Db 687 LSGHLYKINSQYLOKVGIPQSSILSSEFCHYEMEDIDELVSTFKKSVLLRVVDD 746  
OY 835 KCHHAARIRKGSYVQCGIPQSSILSSEFCHYEMEDIDELVSTFKKSVLLRVVDD 793  
Db 747 FYTVNKKDKKFLNLSRGFEKHNFTSLEKTYINENSGIINNTFF 794  
OY 894 FYTVNKKDKKFLNLSRGFEKHNFTSLEKTYINENSGIINNTFF 941  
  
RESULT 5  
ID 013339 PRELIMINARY; PRT; 988 AA.  
AC 013339;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE TELOMERASE REVERSE TRANSCRIPTASE 1.  
GN TRT1.  
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).







```
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TONSIL.
RX MEDLINE: 91305125.
RA DEGUCHI Y., KEHRL J.H.;
RL NUCLEIC ACIDS RES. 19:3742-3742(1991).
CC -1- SIMILARITY: STRONG WITH OTHER "ANTP-TYPE" HOMEOBOX PROTEINS.
DR EMBL: X56537; E30256; ALT_SEQ.
DR PIR: S16681; S16681.
DR PFAM: PF00046; homeobox.
RW DNA-BINDING; NUCLEAR PROTEIN; HOMEOBOX; PHOSPHORYLATION.
FT NON_TER 1
FT DNA_BIND 70 129 HOMEOBOX.
FT DNA_BIND 112 121 H-T-H MOTIF.
SQ SEQUENCE 316 AA; 34713 MW; 1531E766 CRC32;

Query Match 1.9%; Score 136; DB 4; Length 316;
Best Local Similarity 30.9%; Pred. No. 2.67e-04;
Matches 21; Conservative 20; Mismatches 24; Indels 3; Gaps 3;

Db 202 RFWSTASDCSYGRTGAPRGPAT-SPRPSRSPAQDRSPARSAAPGPAASGPGAMTH 260
Oy 223 RAWNHSVREAGV-PLGLPAPGARRRGSASRSLPLKRRPRRG-AAPEPERTPVGGSMAH 280
Db 261 PARPREQA 268
Oy 281 PERTRGPS 288

RESULT 15
ID 062103 PRELIMINARY; PRT; 317 AA.
AC 062103.
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE PROLINE RICH PROTEIN PRECURSOR.
GN PRP.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CD-1: TISSUE-LIVER;
RX MEDLINE: 88273214.
RA ANN D.K., SMITH M.K., CARLSON D.M.;
RL J. BIOL. CHEM. 263:10887-10893(1988).
DR EMBL: M23236; G567232; -.
DR MGD: MGI:97773; PRP.
KW SIGNAL.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 317 PROLINE-RICH PROTEIN.
SQ SEQUENCE 317 AA; 31719 MW; 10C84341 CRC32;

Query Match 1.9%; Score 132; DB 11; Length 317;
Best Local Similarity 26.4%; Pred. No. 9.90e-04;
Matches 39; Conservative 38; Mismatches 64; Indels 7; Gaps 6;

Db 37 SSGQRPYPVNGSQGPPPGPPGPPRP-PQGPPPPGPPPPPPGPPPPPPGPPRP-PQGP 94
Oy 202 AATQARPPPHAS--GPRRLGCE-RAWNHSVREAGVPLGLPAPGARRRGSASRSLPLK 258
Db 95 PPPGPPPPPPGPPPPG-PQRPPPGPPPPGPPPPGPPPPPPGPPPPGPPPPGPPPP 153
Oy 259 RPRGAAPEPERTPVGGSMAHGRGSDRGFCVSPAR-PAEATSLGALSGTRHS 317
Db 154 GGPQRPPPGPPPPGPPPPGPPPPPP 181
Oy 318 HPSVGRQHHAGPPSTSRPPRPWDPCPP 345

Search completed: Fri Dec 18 18:17:04 1998
Job time : 166 secs.
```

**This Page Blank (uspto)**



Query Match	4.48;	Score 97;	DB 27;	Length 1398,
Best Local Similarity	29.38;	Pred. No. 1.85e+01;		

CC one strand of the DNA substrate is cleaved and a nucleic acid  
CC molecule is attached to one or both of the cleaved strands. The



CC receptor designated MEL-1ab, also known as Mel 1-c(beta). The protein differs from previously known receptors by being 65 amino acids shorter and also differs throughout the sequence by 6 amino acids. The protein is encoded by 2 different alleles (T79065-6) which differ in the 3' untranslated region (both sequences encode the same protein), which is thought to affect the half-life of the mRNA. The coding sequence was isolated from cDNA derived from Xenopus skin RNA and amplified using the primers T79067-76. The nucleotide sequence encodes a 7 transmembrane receptor involved in cellular signalling. MEL-1ab has been shown to modulate intracellular cGMP, esp. inhibiting its accumulation induced by an inhibitor of phosphodiesterase, but unlike the MEL-1aa protein, it cannot inhibit adenylyl cyclase activity.

CC Sequence 354 AA;

Query Match 4.1%; Score 90; DB 24; Length 354;

Best Local Similarity 32.1%; Pred. No. 5.64e+01;

Matches 18; Conservative 13; Mismatches 20; Indels 5; Gaps 4;

Db 194 Itvvvvhfvfslsvtfcylfiwvliqvk-rvrgdfkqkltf-talnfltmfv 247

Qy 747 LTEVIASIIKQON--TTC-VKRAVAVQKAAHGKAKKSHVSTLTLQPYMKQFV 799

RESULT 8

ID W23652 standard; Protein; 413 AA.

AC W23652;

DT 09-OCT-1997 (first entry)

DE Recombinant squirrel protease inhibitor 55RS.

KW Protease inhibitor; alpha-1 antitrypsin; HP-55; striped squirrel;

KM hypernatation related protein.

OS Tamias asiaticus.

FH Key Location/Qualifiers

FT peptide 1..24

FT protein /label= Signal

FT region 25..413

FT /label= Protease\_inhibitor\_55RS

FT /label= Reaction\_site 372..381

FT J09157298-A.

PD 17-JUN-1997.

PF 04-DEC-1995; 315692.

PR 04-DEC-1995; JP-315692.

PA (KAND/) KONDO J.

DR (KANA-) 2H KANAGAWA KAGAKU GIUTSU ACAD.

DR WPI; 97-369473/34.

DR N-PSDB: T78180.

PT Squirrel derived protease inhibitor has similar structure to alpha-1

PT antitrypsin - shows selective inhibitory activity against

PT protease(s)

PS Claim 4; Page 8-9; 10pp; Japanese.

CC The present sequence represents a recombinant squirrel protease

CC inhibitor protein 55RS that contains a specifically claimed reaction

CC site, see features table. The protease inhibitor 55RS has inhibitory

CC activity against trypsin, elastase or thrombin. The protease inhibitor

CC has a new structure similar to alpha-1-antitrypsin and shows selective

CC inhibitory activity against proteases. The protease inhibitor cDNA was

CC homologous to cDNA encoding a similar squirrel protein designated

CC hibernation-related protein (HP-55).

CC Sequence 413 AA;

SQ

Query Match 4.1%; Score 90; DB 23; Length 413;

Best Local Similarity 45.8%; Pred. No. 5.64e+01;

Matches 11; Conservative 6; Mismatches 6; Indels 1; Gaps 1;

Db 295 lknr-qtksglyfipkvisgtyd 317

Qy 717 LRVRAQDPPELVYKVDVTGAYD 740

RESULT 9

ID R88409 standard; Protein; 420 AA.

AC R88409;

DT 15-AUG-1996 (first entry)

DE High-affinity melatonin receptor.

KW Melatonin receptor; G-protein-coupled receptor; glycosylation;

KM disulfide bond; ligand binding pocket; phosphorylation;

KW cytochrome-c family heme binding site; melatonin receptor-agonist;

KW melatonin receptor-antagonist; circadian rhythm disorder; jet-lag;

KW day-night cycle disorder; ovulation; reproductive cycle; antibody;

KW animal breeding; puberty; transgenic animal; drug screening.

OS Xenopus laevis.

FH Key Location/Qualifiers

FT domain 1..32

FT /note= "Extracellular domain"

FT modified\_site 5..7

FT /note= "N-glycosylation site"

FT modified\_site 16

FT /note= "Protein-kinase-C phosphorylation site"

FT domain 33..57

FT /note= "Transmembrane region-I"

FT domain 57..68

FT /note= "Intracellular loop"

FT region 67..72

FT /note= "Conserved melatonin receptor motif"

FT domain 69..93

FT /note= "Transmembrane region-II"

FT domain 94..107

FT /note= "Extracellular loop"

FT disulfide\_bond 105..182

FT domain 108..126

FT /note= "Transmembrane region-III"

FT region 125..131

FT /note= "Region used to construct primer"

FT domain 127..151

FT /note= "Intracellular loop"

FT region 132..137

FT /note= "Putative cytochrome-c family heme binding site"

FT modified\_site 137

FT /note= "Protein-kinase-C phosphorylation site"

FT domain 152..171

FT /note= "Transmembrane region-IV"

FT misc\_difference 166

FT /note= "Residue which may form ligand binding pocket"

FT domain 172..193

FT /note= "Extracellular loop"

FT domain 194..220

FT /note= "Transmembrane region-V"

FT misc\_difference 204

FT /note= "Residue which may form ligand binding pocket"

FT domain 221..243

FT /note= "Intracellular loop"

FT domain 244..268

FT /note= "Transmembrane region-VI"

FT region 252..259

FT /note= "Region used to construct primer"

FT misc\_difference 258

FT /note= "Residue which may form ligand binding pocket"

FT domain 269..279

FT /note= "Extracellular loop"

FT domain 280..301

FT /note= "Transmembrane region-VII"

FT region 296..300

FT /note= "Conserved melatonin receptor-1b motif"

FT domain 302..420

FT /note= "Intracellular domain"

FT modified\_site 320

FT /note= "Protein-kinase-C phosphorylation site"

FT modified\_site 328

FT /note= "Protein-kinase-C phosphorylation site"

FT modified\_site 362

FT /note= "Protein-kinase-C phosphorylation site"

FT modified\_site 395

FT /note= "Protein-kinase-C phosphorylation site"

FT modified\_site 418

FT /note= "Protein-kinase-C phosphorylation site"

PN WC9535320-A1.









Query Match	3.98;	Score 86;	DB 8;	Length 1702;
Best Local Similarity	28.68;	Pred. No. 1.05e+02;		
Matches	22;	Conservative	19;	Mismatches 28;
			Indels	8;
			Gaps	7;

Db 1074 gfk-vlydsvsgeellirngkirfv-kikdlfskvdysigekeycilegvealtidd 1131  
|: : | | | : : | : : | | | : : | | |  
OY 651 GLRPIVNM DYVGAR-TRREKRAERLTSRYKALFSLNTERARRP-GIL-GASVGLDD 707  
Db 1132 dqkiwkpvpymthra 1148  
| : : | | |  
OY 708 IHR-AMRT--FVLRYRA 721

Search completed: Fri Dec 18 18:35:58 1998  
Job time : 59 secs.

**THIS PAGE BLANK (USPTO)**

```
RESULT 2
ENTRY 2
TITLE H70127 #type complete
#formal_name Borrelia burgdorferi #common_name Lyme disease
ORGANISM spirochete
DATE 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change
13-Feb-1998

ACCESSIONS
REFERENCE H70127
#authors Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.;
Clayton, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb,
Dodson, R.; Ricketts, R.D.; Richardson, D.; Peterson, J.;
Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;
Vogt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman,
J.; Uterback, T.; Matthey, L.; McDonald, L.; Artlich, P.;
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst,
K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.

#journal Nature (1997) 390:580-586
#title Genomic sequence of a Lyme disease spirochaete, Borrelia
#cross-references M01D:98065943
#accession H70127
#status preliminary: nucleic acid sequence not shown;
translation not shown

#molecule_type DNA
#residues 1-94 #label KLE
#cross-references GB:AE000783; TIGR:BB0224
#experimental_source strain B31
SUMMARY #length 94 #molecular_weight 10659 #checksum 9000

Query Match 4.9%; Score 108; DB 2; Length 94;
Best Local Similarity 26.8%; Pred. No. 3,12e-01;
Matches 19; Conservative 19; Mismatches 30; Indels 3; Gaps 3;

Db 12 LELSFIIISCVRKGIYKNGCINKGISEKELLASVCSNHYDINSINDGIRKAN 71
QY 827 LRFVVF-LRFVCH-HAVAKIKGKSTVQCQIGPQSTL-STLLCSLCYGDMEKRLFRGIRK 883
DB 72 LNLKIKAPLL 82
QY 884 LLRLVDDFL 894

RESULT 3
ENTRY 3
TITLE A64591 #type complete
#formal_name Helicobacter pylori (strain 26695)
ORGANISM 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
DATE 10-Oct-1997

ACCESSIONS
REFERENCE A64591
#authors Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodek, A.;
McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.;
Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Uterback, T.R.;
Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
Fujii, C.; Bowman, C.; Matthey, L.; Mallin, E.; Hayes,
W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
C.M.; Venter, J.C.
#journal Nature (1997) 388:539-547
#title The complete genome sequence of the gastric pathogen
#cross-references M01D:97394467
#accession A64591
#status preliminary: nucleic acid sequence not shown;
translation not shown
#molecule_type DNA

##residues 1-366 ##label TOM
##cross-references GB:AE000571; GB:AE000511; NID:g2313686; PID:g2313689;
TIGR:HP0569
SUMMARY #length 366 #molecular_weight 40574 #checksum 4119

Query Match 4.9%; Score 108; DB 2; Length 366;
Best Local Similarity 25.8%; Pred. No. 3,12e-01;
Matches 16; Conservative 21; Mismatches 22; Indels 3; Gaps 3;

Db 3 LSVGIVGLPNVGS-STENALKTQNAQSANYPCTIEPNKAIYVPPRL-DALAQIVK 60
QY 698 LGSVYGLDIDHRAWTFVLVRADPPPELY-FVKVDTGAYDTIPDRLTEVIASIIK 756
DB 61 PE 62
QY 757 PQ 758

RESULT 4
ENTRY 4
TITLE S49743 #type complete
#formal_name Saccharomyces cerevisiae
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 13-Jan-1995 #sequence_revision 10-Feb-1995 #text_change
21-Nov-1997

ACCESSIONS
REFERENCE S49743
#authors Badcock, K.; Churcher, C.
#submission submitted to the EMBL Data Library, November 1994
#accession S49743
#molecule_type DNA
#residues 1-189 ##label BAD
#cross-references EMBL:246659; NID:g575680; PID:g575683; MIPS:YML036w
GENETICS #map_position 13L
SUMMARY #length 189 #molecular_weight 21771 #checksum 9333

Query Match 4.7%; Score 103; DB 2; Length 189;
Best Local Similarity 26.1%; Pred. No. 1,20e+00;
Matches 23; Conservative 25; Mismatches 35; Indels 5; Gaps 5;

Db 50 GROMYSAYKTLIEV-KYKMKRTRMINSCEVCLSPTSNISDAF-LKFGIKDSSOLICL 107
QY 673 AERTLSRV-KALFSVINTRRARPGLGASVGLDDIHRAWTFVLVRADPPPELYTV 731
DB 108 KFH-TNT-DVDKEDQRLTIMISIVKQE 133
QY 732 KVDVGTAYDTIPDRLTEVIASIIKPN 759

RESULT 5
ENTRY 5
TITLE G64875 #type complete
#formal_name Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
05-Dec-1997

ACCESSIONS
REFERENCE G64875; S22375; A49756
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Colado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references M01D:97426617
#accession G64875
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-891 ##label BLAT
#cross-references GB:AE000225; GB:U00096; NID:g1787523; PID:g1787531;
UMCP:b1276
#experimental_source strain K-12, substrain MG1655
```



```

#authors      Fontaine, J.M.; Rousvoal, S.; Leblanc, C.; Kloareg, B.;
#journal      J. Mol. Biol. (1995) 251:378-389
#title        The mitochondrial LSU rRNA of the brown alga Pyrallela
              littoralis reveals alpha-proteobacterial features and is
              split by four group IIB introns with an atypical phylogeny.
#accession    S58503
#status       preliminary
#molecule_type DNA
#residues     1-318 ##label PON
SUMMARY       #length 318 #checksum 7326

Query Match   4.6%; Score 102; DB 2; Length 318;
Best Local Similarity 46.7%; Pred. NO.1.57e+00;
Matches 14; Conservative 7; Mismatches 8; Indels 1; Gaps 1;

Db 126 SKSY-PSKRGTPGGVLSPLLCNMNLINGLEN 154
QY 844 GKSYVOCGIPGSIPLSLCLCYGDMEN 873

RESULT 9
ENTRY   ORCS      #type complete
TITLE   serine chemoreceptor protein - Escherichia coli
ALTERNATE_NAMES
ORGANISM #formal_name Escherichia coli
DATE     14-Nov-1983 #sequence_revision 13-Feb-1998 #text_change
ACCESSIONS E65250; S56581; A03442; S30281
REFERENCE  A64720
#authors  Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
           Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
           Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
           Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
           Y.
#journal  Science (1997) 277:1453-1462
#title    The complete genome sequence of Escherichia coli K-12.
#cross-references M01D:9742617
#accession E65250
#status    preliminary; nucleic acid sequence not shown;
           translation not shown
#molecule_type DNA
#residues 1-551 ##label BLAT
#cross-references GB:AE000506; GB:U00096; NID:92367377; PID:92367378;
           UMGCP:D4355
REFERENCE  ##experimental_source strain K-12, substrain MG1655
           S56514
#authors  Blattner, F.R.
#journal  Nucleic Acids Res. (1995) 23:2105-2119
#title    Analysis of the Escherichia coli genome VI: DNA sequence of
           the region from 92.8 through 100 minutes.
#cross-references M01D:9534362
#accession S56581
#status    preliminary; nucleic acid sequence not shown;
           translation not shown
#molecule_type DNA
#residues 1-503, 'X', 505-551 ##label BUR
#cross-references EMBL:U14003; NID:91263172; PID:9537197
           ##note the nucleotide sequence was submitted to the EMBL Data
           Library, August 1994
REFERENCE  A03442
#authors  Boyd, A.; Kendall, K.; Simon, M.I.
#journal  Nature (1983) 301:623-626
#title    Structure of the serine chemoreceptor in Escherichia coli.
#cross-references M01D:83141753
#accession A03442
#molecule_type DNA
#residues 1-147, 'S'RSSLSIRPRDIRN', 162-536 ##label BOY
#cross-references GB:V00373; GB:J01718; NID:943217; PID:943218
           S30279
REFERENCE  Roper, D.I.; Pawcett, T.; Cooper, R.A.
#authors  Mol. Gen. Genet. (1993) 237:241-250
#journal

```

```

#title        The Escherichia coli C homoproteocatechuate degradative
              operon: hpc gene order, direction of transcription and
              control of expression.
#accession    S30281
#molecule_type DNA
#residues     1-77 ##label ROP
#cross-references GB:S56952; NID:9298654; PID:9298656
COMMENT       This protein responds to changes in serine concentration in the
              environment, transduces a signal from the outside to the inside
              of the cell, and facilitates sensory adaptation through various
              levels of methylation.
GENETICS
#gene         tsr
#map_position 99 min
CLASSIFICATION #superfamily methyl-accepting chemotaxis protein
KEYWORDS      chemotaxis; methylated amino acid; sensory transduction;
              transmembrane protein
FEATURE
297,311       #modified_site glutamate methyl ester (Glu) (by
              cheB-dependent deamidation and methylation) #status
              experimental\
304,493       #modified_site glutamate methyl ester (Glu) #status
              experimental\
503           #modified_site glutamate methyl ester (Glu) #status
              predicted
SUMMARY       #length 551 #molecular_weight 59442 #checksum 1281

Query Match   4.6%; Score 101; DB 1; Length 551;
Best Local Similarity 22.4%; Pred. NO.2.04e+00;
Matches 22; Conservative 28; Mismatches 44; Indels 4; Gaps 4;

Db 106 ADYFALPRDPROSTRAAEIKRNDYHNMLAEIQLGAGKINEFF-DQ-PTGYODGF 163
QY 737 GADYTRIPD-RLEFVIAIIRKPG-NITCVRRVAVQAAAGHRKAKRSHVSTLIDLPY 794
Db 164 EKQYVAYEQNDRLHDIAVSDNNAYSQAMILVGM 201
QY 795 MRQFVAHLQETSPLRDAVIVQSSSLNEASSGLFDVFL 832

RESULT 10
ENTRY   S51470      #type complete
TITLE   hypothetical protein YLR383w - yeast (Saccharomyces
              cerevisiae)
ALTERNATE_NAMES hypothetical protein L3502.2
ORGANISM #formal_name Saccharomyces cerevisiae
DATE     23-Feb-1995 #sequence_revision 12-May-1995 #text_change
ACCESSIONS S51470; S20170; S20168; S20163
REFERENCE  S51466
#authors  Du, Z.
#journal  Submitted to the EMBL Data Library, December 1994
#title    The sequence of S. cerevisiae cosmid L3502.
#accession S51470
#molecule_type DNA
#residues 1-1114 ##label DUZ
#cross-references EMBL:U19104; NID:9609423; PID:9609425; MIPS:YLR383w
           S15581
REFERENCE  S15581
#authors  Slonimski, P.P.
#journal  Submitted to the EMBL Data Library, October 1987
#accession S20170
#molecule_type DNA
#residues 1-184 ##label SLO
#cross-references EMBL:X05143; NID:94019; PID:94020
           A91079
REFERENCE  Labouesse, M.; Herbert, C.J.; Du Jardin, G.; Slonimski, P.P.
#authors  EMBO J. (1987) 6:713-721
#journal  Three suppressor mutations which cure a mitochondrial RNA
#title    maturase deficiency occur at the same codon in the open
           reading frame of the nuclear NAM2 gene.
#cross-references M01D:87218532
#accession S20168
#status    translation not shown

```

```
##molecule-type DNA
##residues 1-61 #label LAB
##cross-references EMBL:X05143
S18752
#authors Herbert, C.J.; Dujardin, G.; Labouesse, M.; Slonimski, P.P.
#journal Mol. Gen. Genet. (1988) 213:297-309
#title Divergence of the mitochondrial leucyl tRNA synthetase genes
        In two closely related yeasts Saccharomyces cerevisiae and
Saccharomyces douglasii: a paradigm of incipient evolution.
#cross-references MIMD:89039717
#accession S20163
#status not compared with conceptual translation
#molecule-type DNA
##residues 1-62 #label HER
GENETICS
#gene SGD:RHC18
##cross-references SGD:S0004375; MIPS:YLR363w
#map_position 12R
SUMMARY #length 1114 #molecular-weight 128008 #checksum 5560
Query Match          4.6%; Score 101; DB 2; Length 1114;
Best Local Similarity 40.8%; Pred. No. 2,04e+00;
Matches      20; Conservative   12; Mismatches 11; Indels    6; Gaps    5;

Db      45 MTQENSLTDEDCGFRSSDVATRADQNFLEESPSTGIKVIIRNFMCH 93
       111::1::1::1::1::1::1::1::1::1::1::1::1::1::1::1::1::1
Oy      795 MRQEVAH-IDETSPLRDA--VVI-EQSSSLINERSSGLFD-VFLDR-FMCH 837

RESULT  11
ENTRY   GNWVP #type complete
TITLE   genome polyprotein - yellow fever virus (strain Pasteur
         17D-204)
CONTAINS capsid protein C; envelope protein M; major envelope protein
E; nonstructural protein NS1; nonstructural protein NS2a;
nonstructural protein NS2b; nonstructural protein NS3;
nonstructural protein NS4a; nonstructural protein NS4b;
nonstructural protein NS5
#formal_name yellow fever virus
#organism 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change
12-Apr-1996
ORGANISM
DATE     S07757
ACCESSIONS
REFERENCE
#authors Dupuy, A.; Despres, P.; Cahour, A.; Girard, M.; Bouloy, M.
#journal Nucleic Acids Res. (1989) 17:3989
#title Nucleotide sequence comparison of the genome of two 17D-204
yellow fever vaccines.
#cross-references MIMD:89282413
#accession S07757
#status nucleic acid sequence not shown; translation not shown
#molecule-type genomic RNA
##residues 1-3411 #label DUP
##cross-references EMBL:X15062
#note the nucleotide sequence was submitted to the EMBL Data
Library, April 1989, in computer-readable form
CLASSIFICATION #superfamily yellow fever virus genome polyprotein
KEYWORDS capsid protein; envelope protein; glycoprotein; nonstructural
protein; polypeptide; transmembrane protein
FEATURE
2-210 #product capsid protein C #status predicted #label CPC\
105-125 #domain transmembrane #status predicted #label TM1\
211-285 #product envelope protein M #status predicted #label
EPN\
271-289 #domain transmembrane #status predicted #label TM2\
286-778 #product major envelope protein E #status predicted
#label ME\
736-753 #domain transmembrane #status predicted #label TM3\
756-778 #domain transmembrane #status predicted #label TM4\
779-1187 #product nonstructural protein NS1 #status predicted
#label NS1\
1133-1151 #domain transmembrane #status predicted #label TM5\
1160-1179 #domain transmembrane #status predicted #label TM6\
1188-1354 #product nonstructural protein NS2a #status predicted
```

```

1355-1484      #label N2A\
                #product nonstructural protein NS2b #status predicted
1485-2107      #label N2B\
                #product nonstructural protein NS3 #status predicted
2108-2394      #label NS3\
                #product nonstructural protein NS4a #status predicted
2395-2506      #label N4A\
                #product nonstructural protein NS4b #status predicted
2507-3411      #label N4B\
                #product nonstructural protein NS5 #status predicted
                #label NS5\
134,150,172,594,
908,986,1796,2062,
2320,2346,2408,
2467,2720,2734,
2740           #binding_site carbohydrate (Asn) (covalent) #status
                predicted
SUMMARY        #length 3411 #molecular-weight 379528 #checksum 8417

Query Match    4.6%; Score 101; DB 1; Length 3411;
Best Local Similarity 32.3%; Pred. No. 2,04e+00;
Matches 21; Conservative 14; Mismatches 27; Indels 3; Gaps 3;

Db   1334 GUTOPFLG-LCAFLARIRGRSIPVNEALAAAGLVYLAGLAFQEMEN-FLGIPIANGCL 1391
     || : || : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy   826 GLEDFLARPMCHH-AVRIGKSYVQCQIGPSILSTLLCSLCGYDMENKLFAGRIRDGL 884

Db   1392 LMLVY 1396
     | : | |
Qy   885 LRLRV 889

RESULT 12
ENTRY    GNMVY          #type complete
TITLE    genome polyprotein - yellow fever virus (strain 17D)
CONTAINS E; capsid protein C; envelope protein M; major envelope protein
         E; nonstructural protein NS1; nonstructural protein NS2a;
         nonstructural protein NS2b; nonstructural protein NS3;
         nonstructural protein NS4a; nonstructural protein NS4b;
         nonstructural protein NS5
ORGANISM #formal_name yellow fever virus
DATE      27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change
          12-Apr-1996
ACCESSION A03914
REFERENCE A03914
AUTHORS   Rice, C.M.; Lenches, E.M.; Eddy, S.R.; Shin, S.J.; Sheets,
          R.L.; Strauss, J.H.
          Science (1985) 229:726-733
          Nucleotide sequence of yellow fever virus: implications for
          flavivirus gene expression and evolution.
#cross-references MDID:85272570
#accession A03914
FEATURES ##molecule_type genomic RNA
          ##residues 1-3411 ##label RIC
CLASSIFICATION #superfamily yellow fever virus genome polyprotein
KEYWORDS       capsid protein; envelope protein; glycoprotein; nonstructural
               protein; polyprotein; transmembrane protein
FEATURE
2-210         #product capsid protein C #status predicted #label CPC\
211-285       #product envelope protein M #status predicted #label
              EPM\
249-269       #domain transmembrane #status predicted #label TM1\
271-285       #domain transmembrane #status predicted #label TM2\
286-778       #product major envelope protein E #status predicted
              #label MEEL
740-753       #domain transmembrane #status predicted #label TM3\
755-778       #domain transmembrane #status predicted #label TM4\
779-1187      #product nonstructural protein NS1 #status predicted
              #label NS1\
1159-1180     #domain transmembrane #status predicted #label TM5\
1188-1354     #product nonstructural protein NS2a #status predicted
              #label N2A\
1355-1484     #product nonstructural protein NS2b #status predicted

```



\*\*\*\*\*  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Blocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Fri Dec 18 18:37:17 1998; Maspar time 9.66 Seconds  
 Tabular output not generated. 836.466 Million cell updates/sec

Title: >US-08-951-733-14  
 Description: (640-940) from US08951733.pep (2 of 2)  
 Perfect Score: 2214  
 Sequence: 1 TSKLRFIPKPDGLRPIVND.....NKRKYVNPVDEALGTA 301

Scoring table:  
 PAM 150  
 Gap 11

Searched: 74019 seqs, 26840295 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: swiss-prot35  
 1:swissprot

Statistics: Mean 49.063; Variance 85.433; scale 0.574

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	103	4.7	181	1	YMD6_YEAST	2.00e-01
2	105	4.7	890	1	ACOL1_ECOLI	1.06e-01
3	103	4.7	891	1	ACON1_ECOLI	2.00e-01
4	102	4.6	205	1	YOO1_BPBP1	2.75e-01
5	101	4.6	551	1	MCPL1_ECOLI	3.76e-01
6	101	4.6	1114	1	RH18_YEAST	3.76e-01
7	101	4.6	3411	1	POLG_YEAV1	3.76e-01
8	101	4.6	3411	1	POLG_YEAV2	3.76e-01
9	99	4.5	446	1	MOXY_PARDE	5.14e-01
10	100	4.5	1030	1	YEO2_SCHPO	5.14e-01
11	98	4.4	586	1	YCE1_YEAST	9.52e-01
12	98	4.4	906	1	ACON1_ECOLI	9.52e-01
13	97	4.4	1238	1	PTPJ_MOUSE	1.29e+00
14	96	4.3	318	1	VP23_HSV11	1.74e+00
15	95	4.3	362	1	ALFC_CARAU	2.35e+00
16	96	4.3	417	1	ASPG_METJA	1.74e+00
17	96	4.3	789	1	ATZM_YEAST	1.74e+00
18	92	4.2	212	1	Y1010_MYCN	4.24e+00
19	93	4.2	261	1	YGF6_ECOLI	4.24e+00
20	92	4.2	455	1	YAO6_SCHPO	5.67e+00
21	92	4.2	647	1	UUP1_HAEIN	5.67e+00
22	94	4.2	781	1	PARC_MYGE	3.16e+00
23	94	4.2	901	1	PHSG_YEAST	3.16e+00

24	92	4.2	1117	1	YDVG_SCHPO	PUTATIVE HELICASE GCF1	5.67e+00
25	91	4.1	124	1	RR12_EUGER	CHLOROPLAST 30S RIBOSO	7.56e+00
26	91	4.1	164	1	PHEA_PREDI	C-PHYCOERYTHRIN ALPHA	1.00e+01
27	90	4.1	297	1	YXG_BACSD	HYPOTHETICAL 32.5 KD P	1.00e+01
28	90	4.1	309	1	VIRB_SHIFT	VIRULENCE REGULON TRAN	1.00e+01
29	90	4.1	357	1	ALF_SPIOL	FRUCTOSE-BISPHOSPHATE	1.00e+01
30	91	4.1	366	1	YXG_ECOLI	HYPOTHETICAL 40.4 KD P	1.00e+01
31	90	4.1	420	1	MLC_XENLA	MELATONIN RECEPTOR TYP	1.00e+01
32	90	4.1	428	1	DD11_YEAST	DNA-DAMAGE INDUCIBLE P	1.00e+01
33	90	4.1	433	1	FUSA_BURCE	FUSARIC ACID RESISTANC	1.00e+01
34	90	4.1	445	1	LAEB_VIBPA	FLAGELLAR HOOK-ASSOCIA	1.00e+01
35	90	4.1	452	1	YHA2_YEAST	HYPOTHETICAL 51.2 KD P	1.00e+01
36	91	4.1	609	1	UVRC_YEAST	EXCINCLASE ABC SUBUN	7.56e+00
37	91	4.1	695	1	PIP6_BOVIN	6-PHOSPHORUTOKINASE,	7.56e+00
38	91	4.1	780	1	K6PL_MOUSE	SEX-DETERMINING TRANS	1.00e+01
39	90	4.1	1475	1	TRA2_CABEL	PROTEIN KINASE GCN2 (E	7.56e+00
40	91	4.1	1590	1	GCN2_YEAST	CHLOROPLAST 30S RIBOSO	1.33e+01
41	89	4.0	124	1	RR12_HORVU	TRAM PROTEIN.	1.33e+01
42	89	4.0	127	1	TRM8_ECOLI	CHLOROPLAST 30S RIBOSO	1.33e+01
43	89	4.0	133	1	RR12_CHLRE	TRANSCRIPTIONAL REGULA	1.33e+01
44	89	4.0	532	1	RTRC_ECOLI	NITRITE REDUCTASE (NAD	1.33e+01
45	89	4.0	957	1	NIRB_KLEBN		

## ALIGNMENTS

RESULT	ID	YMD6_YEAST	STANDARD:	PRT:	181 AA.
AC	Q03705	01-NOV-1997 (REL. 35, CREATED)			
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)				
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)				
DE	HYPOTHETICAL 20.7 KD PROTEIN IN CAT2-AND1 INTERGENIC REGION.				
GN	YMD036W.				
OS	SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).				
OC	EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRATIN-S288C / AB972;				
RA	BADCOCK K., CHURCHER C., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;				
RL	SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.				
DR	EMBL: Z46559; G575683; ALT_SEQ.				
KW	HYPOTHETICAL PROTEIN.				
QY	SEQUENCE 181 AA; 20662 MW; AC111C34 CRC32;				
DB	50 GEOMYSALYKTLIEV-KYMKRTNLNLSCEVLCSPNSISDAE-LKFGIKDSSOLICL 107				
QY	673 AERLTISRV-KALFESVLYNERARRRGLGASVLDGDDIHRAMRTFLVLRADDPPELYFV 731				
DB	108 KFH-TNT-DVDYKREQLRTIMTSYKGOE 133				
QY	732 KVDVTGAYDITIPDRLEVIASITIKPON 759				
RESULT	2				
ID	ACOL1_ECOLI	STANDARD:	PRT:	890 AA.	
AC	P25516; P78060; P78148;				
DT	01-MAY-1992 (REL. 22, CREATED)				
DT	01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)				
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)				
DE	ACONITATE HYDRATASE 1 (EC 4.2.1.3) (CITRATE HYDRO-LYASE 1)				
GN	ACONITATE 1).				
GN	ACNA OR ACN.				
OS	ESCHERICHIA COLI.				
OC	PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;				
OC	ENTEROBACTERIACEAE.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 92174916.				

RA PRODROMOU C., ARTYMIUK P.J., GUEST J.R.;  
RL EUR. J. BIOCHEM. 204:599-609(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-R12 / MG1655;  
RX BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;  
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-R12;  
RA AIBA H., BABI T., FUJITA K., HAYASHI K., HONJO A., HORIUCHI T.,  
RA IKEMOTO K., INADA T., ISONO K., ISONO S., ITOH T., KANAI K.,  
RA KASAI H., KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M.,  
RA KITAHARA M., MAKINO K., MASUDA S., MIKI T., MIZOBUCHI K., MORI H.,  
RA KOTAKURA K., NAKAMURA Y., NASHIMOTO K., NISHIO Y., OSHIMA T.,  
RA SAITO N., SAMPEI G., SEKI Y., TAGAMI H., TAKEMOTO K., WADA C.,  
RA YAMAMOTO Y., YANO M.;  
RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [4]  
RP SEQUENCE OF 1-18.  
RX MEDLINE: 92148368.  
RA PRODROMOU C., HAYNES M.J., GUEST J.R.;  
RL J. GEN. MICROBIOL. 137:2505-2515(1991).  
CC -1- FUNCTION: MAY HAVE AN IRON-RESPONSIVE REGULATORY FUNCTION.  
CC -1- CATALYTIC ACTIVITY: CITRATE - CIS-ACONITATE + H(2)O.  
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.  
CC -1- SUBUNIT: MONOMER.  
CC -1- ACONITATE HAS AN ACTIVE (4FE-4S) AND AN INACTIVE (3FE-4S) FORMS.  
CC THE ACTIVE (4FE-4S) CLUSTER IS PART OF THE CATALYTIC SITE THAT  
CC INTERCONVERTS CITRATE, CIS-ACONITATE, AND ISOCITRATE.  
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.  
DR EMBL: AE000225; G1787531; -;  
DR EMBL: D90765; G1742087; -;  
DR EMBL: D90766; G1742090; -;  
DR PIR: S22375; S22375.  
DR ECOGENE: EG11325; ACNA.  
DR PROSITE: PS00450; ACONITASE\_1; 1.  
DR PROSITE: PS01244; ACONITASE\_2; 1.  
KW LYASE: TRICARBOXYLIC ACID CYCLE; IRON-SULFUR; 4FE-4S.  
FT INIT\_MET 0  
FT METAL 434 434  
FT METAL 500 500  
FT METAL 503 503  
FT METAL 503 503  
FT CONFLICT 521 521  
FT SEQUENCE 890 AA; 97516 MW; 276673F6 CRC32;  
SQ  
Query Match 4.7%; Score 105; DB 1; Length 890;  
Best local Similarity 21.9%; Pred. No. 1,06e-01;  
Matches 14; Conservative 22; Mismatches 25; Indels 3; Gaps 3;  
Db 754 MRYKQEQTPIAVLAGKEVGGSSR-DMAAKGPRLLG-IRVAVIE-SFERIHRNSLIGMGI 810  
Qy 643 LRPIKPDGRLPIVNDYVVGARTFRERKARLTSSVKALFSLNTERARRRGLGASV 702  
Db 811 LPLE 814  
Qy 703 LGLD 706  
RESULT 3  
ID ACON\_LEGPN STANDARD; PRT; 891 AA.  
AC P37032;  
DT 01-JUN-1994 (REL. 29, CREATED)  
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)  
DE ACONITATE HYDRAITASE (EC 4.2.1.3) (CITRATE HYDRO-LYASE) (ACONITASE)  
DE (MAJOR IRON-CONTAINING PROTEIN) (MICP) (IP210).  
GN ACN.  
OS LEGIONELLA PNEUMOPHILA.

OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;  
OC LEGIONELLACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PHILADELPHIA 1;  
RX MEDLINE: 93374864.  
RA MENGAUD J.M., HORWITZ M.A.;  
RL J. BACTERIOL. 175:5666-5676(1993).  
CC -1- CATALYTIC ACTIVITY: CITRATE - CIS-ACONITATE + H(2)O.  
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.  
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
CC -1- ACONITASE HAS AN ACTIVE (4FE-4S) AND AN INACTIVE (3FE-4S) FORMS.  
CC THE ACTIVE (4FE-4S) CLUSTER IS PART OF THE CATALYTIC SITE THAT  
CC INTERCONVERTS CITRATE, CIS-ACONITATE, AND ISOCITRATE.  
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.  
DR EMBL: 122081; G348945; -;  
DR PIR: B48642; B48642.  
DR HSSP: P20004; IACO.  
DR PROSITE: PS00450; ACONITASE\_1; 1.  
DR PROSITE: PS01244; ACONITASE\_2; 1.  
KW LYASE: TRICARBOXYLIC ACID CYCLE; IRON-SULFUR; 4FE-4S.  
FT METAL 435 435  
FT METAL 501 501  
FT METAL 504 504  
FT METAL 504 504  
FT SEQUENCE 891 AA; 98209 MW; E6230AC5 CRC32;  
SQ  
Query Match 4.7%; Score 103; DB 1; Length 891;  
Best local Similarity 25.0%; Pred. No. 2,00e-01;  
Matches 16; Conservative 22; Mismatches 23; Indels 3; Gaps 3;  
Db 755 MRYQEQDPLVLAGKEVGTSSR-DMAAKGTLLG-VKAVITE-SFERIHRNSLIGMGI 811  
Qy 643 LRPIKPDGRLPIVNDYVVGARTFRERKARLTSSVKALFSLNTERARRRGLGASV 702  
Db 812 LPLQ 815  
Qy 703 LGLD 706  
RESULT 4  
ID Y001\_BPHPI STANDARD; PRT; 205 AA.  
AC P51700;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 23.2 KD PROTEIN IN INT-CL INTERGENIC REGION (ORF1).  
DE BACTERIOPHAGE HPI.  
OS VIRIDAE; DS-DNA NONENVELOPED VIRUSES; MYOVIRIDAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-HP1C1;  
RX MEDLINE: 95089704.  
RA ESPOSITO D., SCOCOA J.J.;  
RL MOL. MICROBIOL. 13:685-695(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-HP1C1;  
RX MEDLINE: 96279738.  
RA ESPOSITO D., FITZMAURICE W.P., BENJAMIN R.C., GOODMAN S.D.,  
RA WALDMAN A.S., SCOCOA J.J.;  
RL NUCLEIC ACIDS RES. 24:2360-2368(1996).  
DR EMBL: U24159; G459176; -;  
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.  
FT TRANSMEM 12 69  
FT TRANSMEM 49 32  
FT SEQUENCE 205 AA; 23260 MW; 821C7A9F CRC32;  
SQ  
Query Match 4.6%; Score 102; DB 1; Length 205;  
Best local Similarity 22.1%; Pred. No. 2,75e-01;  
Matches 33; Conservative 49; Mismatches 50; Indels 17; Gaps 17;



RL	EMBO J.6:713-721(1987).
CC	-I- FUNCTION: ACTS IN A DNA REPAIR PATHWAY FOR REMOVAL OF UV-INDUCED
CC	DNA DAMAGE THAT IS DISTINCT FROM CLASSICAL NUCLEOTIDE EXCISION
CC	REPAIR AND IN REPAIR OF IONIZING RADIATION DAMAGE (BY SIMILARITY).
CC	-I- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC	-I- SIMILARITY: BELONGS TO THE SMC FAMILY.
DR	EMBL; X80930; E198113; -
DR	EMBL; U19104; G609425; -
DR	EMBL; X05143; G4020; -
DR	SGG; L0004130; RRC18.
KW	DNA REPAIR; ATP-BINDING; COILED COIL; NUCLEAR PROTEIN.
FT	DOMAIN 35 39 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT	NP BIND 109 116 ATP (POTENTIAL).
FT	DOMAIN 259 529 COILED COIL (POTENTIAL).
FT	DOMAIN 656 969 COILED COIL (POTENTIAL).
SO	SEQUENCE 1114 AA; 128008 MW; BC94F87E CRC32;
Query Match	4.6%; Score 101; DB 1; Length 1114;
Besst Local Similarity	40.8%; Pred.No.3,76e-01;
Matches 20; Conservative	12; Mismatches 11; Indels 6; Gaps 5.
Db	45 MTQENSLTDESDGERSSSVATADQDNFLSESPSGYKKYLNNMCH 93
Oy	795 MKGFAPHA-LOETSPLRDA--VVI-EQSSSLNKAASSGLFD-VFLR-FMCH 837
RESULT	7
ID	POLG.YEYVL STANDARD; PRT; 3411 AA.
AC	P03314; O42028;
DT	21-JUL-1986 (REL. 01, CREATED)
DT	21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DI	15-JUL-1988 (REL. 36, LAST ANNOTATION UPDATE)
DE	GENOME POLYPEPTIDE (CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
DE	PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL
DE	RNA POLYMERASE (EC 2.7.7.48) (NS3).
OS	YELLOW FEVER VIRUS (STRAIN 11D).
OC	VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; FLAVIVIRIDAE;
OC	FLAVIVIRUSES.
RA	[1]
RX	SEQUENCE FROM N.A.
RP	MEDLINE: 85272570.
RY	RICE C.M., LENCHES E.M., EDDY S.R., SHIN S.J., SHEETS R.L.,
SA	STRAUSS J.H.,
SC	SCIENCE 229:726-733(1985).
CC	-I- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC	HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC	NS3 AND NS5 MAY PLAY A ROLE IN THE VITAL RNA REPLICATION.
CC	-I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC	LIPIDOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC	PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC	PROTEIN C AND MRNA.
CC	EMBL; X03700; G59339; -
DR	EMBL; K02749; G336193; -
DR	PIR; A03914; GNMVY.
KW	POLYPEPTIDE; GLYCOPROTEIN; RNA-DIRECTED RNA POLYMERASE; CORE PROTEIN;
KW	COAT PROTEIN; ENVELOPE PROTEIN; HELICASE; ATP-BINDING; TRANSMEMBRANE;
KW	NONSTRUCTURAL PROTEIN.
FT	INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT	CELLULAR AMINOPEPTIDASE.
FT	CAPSID PROTEIN C.
FT	CHAIN 2 121 ENVELOPE GLYCOPROTEIN M.
FT	PROPEP 122 210 MAJOR ENVELOPE PROTEIN E.
FT	CHAIN 211 285 NONSTRUCTURAL PROTEIN NS1.
FT	CHAIN 286 778 NONSTRUCTURAL PROTEIN NS2B.
FT	CHAIN 779 1130 HELICASE (NS3).
FT	CHAIN 1131 1354 NONSTRUCTURAL PROTEIN NS4A.
FT	CHAIN 1355 1484 NONSTRUCTURAL PROTEIN NS4B.
FT	CHAIN 1485 2107 RNA-DIRECTED RNA POLYMERASE (NS5).
FT	CHAIN 2108 2256 POTENTIAL.
FT	CHAIN 2257 2506
FT	CHAIN 2507 3411
FT	TRANSMEM 249 269
FT	TRANSMEM 271 285

FT	TRANSMEM	740	753	POTENTIAL.
FT	TRANSMEM	755	778	POTENTIAL.
FT	TRANSMEM	1159	1180	POTENTIAL.
FT	DOMAIN	383	396	INVOLVED IN FUSION.
FT	NP_BIND	1682	1689	ATP (POTENTIAL).
FT	SITE	1773	1776	DEAD BOX.
FT	DISULFID	288	315	BY SIMILARITY.
FT	DISULFID	345	401	BY SIMILARITY.
FT	DISULFID	359	390	BY SIMILARITY.
FT	DISULFID	377	406	BY SIMILARITY.
FT	DISULFID	467	568	BY SIMILARITY.
FT	DISULFID	585	615	BY SIMILARITY.
FT	CARBOHYD	134	134	POTENTIAL.
FT	CARBOHYD	150	150	POTENTIAL.
FT	CARBOHYD	908	908	POTENTIAL.
FT	CARBOHYD	986	986	POTENTIAL.
FT	CARBOHYD	2320	2320	POTENTIAL.
FT	CARBOHYD	2346	2346	POTENTIAL.
FT	CARBOHYD	2467	2467	POTENTIAL.
SO	SEQUENCE	3411 AA:	379512 MW:	E85D316D CRC32;
Query Match				
Best Local Similarity	4.6%;	Score 101;	DB 1;	Length 3411;
Matches	21;	Conservative	14;	Mismatches 27; Indels 3; Gaps 3
Db	1334	GLTGFELG-LCAFLATRIEGRSRIPVNEALAAAGLVGLAGLFOEMEN-FLGPIAVGGL	1391	
Oy	826	GLGFVFLRFMCNH-AVRIGRSKYQCQGISILTLCSLCYGDENKFLAGIRRDGL	884	
Db	1392	LMMLV	1396	
Oy	885	LRLEV	889	
RESULT 8				
ID	POLG_YEFV2	STANDARD;	PRT;	3411 AA.
AC	P19901;			
DT	01-FEB-1991 (REL. 17, CREATED)			
DT	01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)			
DT	01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)			
DE	GENOME POLYPROTEIN (CONTAINS: CAPSID PROTEIN C (CORE PROTEIN)); MATRIX PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL PROTEINS NS1, NS2A, NS2B, NS4A AND NS4B; HELICASE (NS3); RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (NS5)).			
DE	RNA POLYMERASE (EC 2.7.7.48) (NS5)).			
OS	YELLOW FEVER VIRUS (STRAIN PASTER 17D-204).			
OC	VIRIDAE: SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; FLAVIVIRIDAE; FLAVIVIRUSES.			
CC	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE; 89282413.			
RL	NUCLEIC ACIDS RES. 17:3989-3989(1989).			
CC	-I- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.			
CC	NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.			
CC	-I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA.			
CC	EMBL; X15062; -: NOT_ANNOTATED_CDS.			
DR	PIR; S07757; GNAVYP.			
KW	CAPSPROTEIN; GLYCOPROTEIN; RNA-DIRECTED RNA POLYMERASE; CORE PROTEIN; MAJOR ENVELOPE PROTEIN; HELICASE; ATP-BINDING; TRANSMEMBRANE; NONSTRUCTURAL PROTEIN.			
KW	NONSTRUCTURAL PROTEIN.			
FT	INIT_MET	1	1	REMOVED FROM CAPSID PROTEIN C BY THE CELLULAR AMINOPEPTIDASE.
FT	CHAIN	2	121	CAPSID PROTEIN C.
FT	PROPEP	122	210	ENVELOPE GLYCOPROTEIN M.
FT	CHAIN	211	285	MAJOR ENVELOPE PROTEIN E.
FT	CHAIN	286	778	NONSTRUCTURAL PROTEIN NS1.
FT	CHAIN	779	1130	NONSTRUCTURAL PROTEIN NS2A.
FT	CHAIN	1131	1354	NONSTRUCTURAL PROTEIN NS2B.
FT	CHAIN	1355	1484	NONSTRUCTURAL PROTEIN NS2B.

Query Match	4.68;	Score 101;	DB 1;	Length 3411;
Best Local Similarity	32.38;	Pred. No. 3.76e-01;		
Matches	21;	Conservative	14;	Mismatches 27; Indels 3; Gaps 3;

RESULT	9	
ID	MOXY	PARDE
	STANDARD;	PRT;
		446 AA

OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI  
OC UNCERTAIN.  
OC

CC EXPRESSION. MAY FUNCTION AS A MEMBRANE-ASSOCIATED PROTEIN KINASE  
CC THAT PHOSPHORYLATES MOXX IN RESPONSE TO ENVIRONMENTAL SIGNALS.

UN 6'IN/ 0320/1, 0320/1.  
 KW SENSORY TRANSDUCTION; TRANSFERASE; KINASE; PHOSPHORYLATION  
 KW TRANSMEMBRANE; INNER MEMBRANE; METHANOL UTILIZATION.

FT	TRANSMEM	101	121	POTENTIAL.
FT	TRANSMEM	144	164	POTENTIAL.
SQ	SEQUENCE	446 AA;	48236 MM;	D59FB521 CRC32

QY 676 LTSRVKALFVSUNYERARRPGLLGASTVLGDDIHRAMRTFVLRVADDPPELFFVKVDV 735

RESULT	10	STANDARD;	PRT;	1030 AA
ID	YE02_SCHPO			
1	0370			

RA SKELTON J., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WOOD V.  
RC STRAIN=972;  
RF SEQUENCE FROM N.A.

Query match	4.35;	Score 100;	DB 1;	Length 1050;
Best Local Similarity	22.78;	Pred. No. 5.14e-01;		
Matches	22;	Conservative	25;	Mismatches 47;
				Indels 3;
				Gaps 3

RESULT	11	
ID	YCFL_YEAST	STANDARD; PRT; 586 AA

RA FULLER L.J., KELLY A., LEWIS C., MCKEE R.A., PEARSON B.M.;  
RL SUBMITTED (MAR-1992) TO EMBL/GENBANK/DBJ DATA BANKS.

DR PIR; S19381; S19381.  
 KW HYPOTHETICAL PROTEIN.  
 SQ SEQUENCE 586 AA; 65152 MW; 56434E7A CRC32

Query Match 4.48; Score 98; DB 1; Length 586;



OY 744 QDR-LTEVIASIKPQNTYCVRRVAVOKAAHGHVR-KAFKSHVSTLDDLP 793

RESULT 14  
ID VP23\_HSV11 STANDARD; PRT; 318 AA.  
AC P10202;  
DT 01-MAR-1989 (REL. 10, CREATED)  
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
DT 01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)  
DE CAPSID PROTEIN VP23.  
GN UL18.  
OS HERPES SIMPLEX VIRUS (TYPE 1 / STRAIN 17).  
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; ALPHAHERPESVIRINAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MCEEOCH D.J., DALRYMPLE M.A., DAVIDSON A.J., DOLAN A., FRAME M.C.,  
RA MCNAB D., PERRY L.J., SCOTT J.E., TAYLOR P.,  
RL J. GEN. VIROL. 69:1531-1574(1988).  
RN [2]  
RP FUNCTION.  
RX MEDLINE; 90264854.  
RA RIXON F.J., DAVIDSON M.D., DAVIDSON A.J.;  
RL J. GEN. VIROL. 71:1211-1214(1990).  
RN [3]  
RP SEQUENCE OF 1-13.  
RX MEDLINE; 93019027.  
RA DAVIDSON M.D., RIXON F.J., DAVIDSON A.J.;  
RL J. GEN. VIROL. 73:2709-2713(1992).  
CC -1- FUNCTION: MAY BE A COMPONENT OF TRIMERIC STRUCTURES LINKING  
CC ADJACENT CAPSOMERES AT THE CAPSID SURFACE.  
CC -1- SIMILARITY: TO OTHER HERPESVIRUSES CAPSID PROTEIN VP23.  
DR EMBL; D10879; G221739; -  
DR EMBL; X14112; G59518; -  
DR EMBL; X14112; E312332; -  
DR PIR; I30083; WMBET8.  
KW COAT PROTEIN  
SQ SEQUENCE 318 AA; 34270 MW; FFEBO01F CRC32;

Query Match 4.3%; Score 96; DB 1; Length 318;  
Best Local Similarity 34.2%; Pred. No. 1.74e+00;  
Matches 27; Conservative 16; Mismatches 28; Indels 8; Gaps 7;

Db 137 LRFVPLPSPLAREIVAR-LV-ARGIRDLPSPNPGGLPDL-NVLYYN-GSRISLL-A 190  
OY 643 LRF-IPKRDGL-RPIVNDYVGARTFRERKARLTSRVKALFSLVYERARRPGLLGA 700  
Db 191 DVQOLGPVNAELRSVLYNM 209  
OY 701 SVLGDDIHRAWRTFVLRV 719

RESULT 15  
ID ALFC\_CARAU STANDARD; PRT; 362 AA.  
AC P53448;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE FRUCTOSE-BISPHOSPHATE ALDOLASE C (EC 4.1.2.13) (BRAIN).  
GN ALDOC.  
OS CARASSIUS AURATUS (GOLDFISH).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; PISCES; GNATHOSTOMATA;  
OC OSTEIFIHTHES; ACTINOPTERYGII; CYPRINIFORMES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN;  
RA REICHERT T.Z., WILLIAMS M.D., CALLARD G.V., TOLAN D.R.;  
RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- CATALYTIC ACTIVITY: D-FRUCTOSE 1,6-BISPHOSPHATE - GLYCERONE-  
CC PHOSPHATE + D-GLYCERALDEHYDE 3-PHOSPHATE.  
CC -1- PATHWAY: SIXTH STEP IN GLYCOLYSIS.  
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).

CC -1- IN VERTEBRATES, THREE FORMS OF THIS UBQUITOUS GLYCOLYTIC ENZYME  
CC ARE FOUND. ALDOLASE A IN MUSCLE, ALDOLASE B IN LIVER, & ALDOLASE  
CC C IN BRAIN.  
CC -1- SIMILARITY: BELONGS TO CLASS I FRUCTOSE-BISPHOSPHATE ALDOLASE  
CC FAMILY.  
DR EMBL; U36777; G1143277; -  
DR PROSITE; PS00158; ALDOLASE CLASS I; 1.  
KW LYSASE; SCHIFF BASE; GLYCOLYSIS; MULTIGENE FAMILY.  
FT INT\_MET 0 0  
FT BINDING 55 55  
FT BINDING 146 146  
FT BINDING 229 229  
FT ACT\_SITE 362 362  
SQ SEQUENCE 362 AA; 39343 MW; 24B98F89 CRC32;

Query Match 4.3%; Score 95; DB 1; Length 362;  
Best Local Similarity 25.4%; Pred. No. 2.35e+00;  
Matches 16; Conservative 22; Mismatches 20; Indels 5; Gaps 5;

Db 147 WRS-VLKI-SETSPSELAMENAVLARYASICOQNGIPIVEBELLPGDHDLCROQV 204  
OY 712 WRTFLRVRAQDPPELFEV-KVDVGYADTI-PODLTEVIASIKPQNTYCVRRVA-V 768  
Db 205 TEK 207  
OY 769 VOK 771

Search completed: Fri Dec 18 18:37:42 1998  
Job time : 25 secs.

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

\*\*\*\*\*  
WIDEORIT (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSrch\_p protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Dec 18 18:39:23 1998; MasPar time 3.38 Seconds  
Tabular output not generated. 629.739 Million cell updates/sec

Title: >US-08-951-733-14  
Description: (640-940) from US08951733.pep (2 of 2)  
Perfect Score: 2214  
Sequence: 1 TSTRIRFPKPDGLRPYVNM.....NLKRTVVFVEDEALGCTA 301  
Scoring table: PAM 150  
Gap 11

Searched: 77309 seqs, 7078906 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5\_COMB 2:PCT9\_COMB 3:backfiles1

Statistics: Mean 31.637; Variance 143.394; scale 0.221

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	97	4.4	903	1	US-08-750-Sequence 1, Applicatio	6.92e+00
2	97	4.4	1398	1	US-08-750-Sequence 9, Applicatio	6.92e+00
3	96	4.3	785	1	US-08-526-Sequence 4, Applicatio	8.14e+00
4	90	4.1	433	1	US-07-661-Sequence 2, Applicatio	2.13e+01
5	85	3.8	880	2	PCT-US94-0-Sequence 5, Applicatio	4.65e+01
6	85	3.8	890	1	US-08-145-Sequence 5, Applicatio	4.65e+01
7	82	3.7	95	1	US-08-202-Sequence 23, Applicatio	7.38e+01
8	81	3.7	858	2	PCT-US93-0-Sequence 2, Applicatio	8.59e+01
9	81	3.7	1178	1	US-08-199-Sequence 2, Applicatio	8.59e+01
10	81	3.7	1178	2	PCT-US95-0-Sequence 2, Applicatio	8.59e+01
11	81	3.7	1713	2	PCT-US94-1-Sequence 24, Applicatio	8.59e+01
12	79	3.6	567	1	US-08-457-Sequence 4, Applicatio	1.16e+02
13	79	3.6	567	1	US-08-458-Sequence 4, Applicatio	1.16e+02
14	79	3.6	567	1	US-08-258-Sequence 4, Applicatio	1.16e+02
15	79	3.6	567	1	US-08-457-Sequence 4, Applicatio	1.16e+02
16	79	3.6	567	1	US-08-457-Sequence 4, Applicatio	1.16e+02
17	79	3.6	567	1	US-08-456-Sequence 4, Applicatio	1.16e+02
18	79	3.6	604	1	US-08-328-Sequence 12, Applicatio	1.16e+02
19	80	3.6	791	1	US-08-394-Sequence 2, Applicatio	1.00e+02
20	80	3.6	890	1	US-08-472-Sequence 8, Applicatio	1.00e+02
21	79	3.6	921	1	US-08-818-Sequence 2, Applicatio	1.16e+02
22	79	3.6	921	1	US-08-396-Sequence 2, Applicatio	1.16e+02
23	80	3.6	1247	1	US-08-472-Sequence 10, Applicatio	1.00e+02

RESULT	ID	US-08-750-532-1	STANDARD;	PRT;	903 AA.
24	80	3.6	3218	1	US-08-764-Sequence 27, Applicati
25	77	3.5	136	3	5212286-4 Patent No. 5212286.
26	77	3.5	146	1	US-08-453-Sequence 38, Applicati
27	77	3.5	146	1	US-08-268-Sequence 38, Applicati
28	77	3.5	151	3	5212286-2 Patent No. 5212286.
29	78	3.5	229	2	PCT-US96-0 Sequence 62, Applicati
30	78	3.5	229	2	PCT-US96-0 Sequence 13, Applicatio
31	77	3.5	296	2	PCT-US95-0 Sequence 2, Applicatio
32	77	3.5	405	1	US-07-994-Sequence 2, Applicatio
33	77	3.5	405	1	US-08-421-Sequence 2, Applicatio
34	77	3.5	405	1	US-07-829-Sequence 2, Applicatio
35	77	3.5	419	1	US-08-553-Sequence 3, Applicatio
36	77	3.5	440	1	US-08-459-Sequence 3, Applicatio
37	77	3.5	619	1	US-07-762-Sequence 2, Applicatio
38	77	3.5	644	2	PCT-US93-0 Sequence 21, Applicati
39	77	3.5	647	2	PCT-US93-0 Sequence 11, Applicati
40	77	3.5	647	2	PCT-US93-0 Sequence 23, Applicati
41	77	3.5	662	1	US-08-261-Sequence 7, Applicatio
42	78	3.5	699	2	PCT-US94-0 Sequence 39, Applicatio
43	77	3.5	735	2	PCT-US93-0 Sequence 13, Applicati
44	78	3.5	1019	1	US-08-271-Sequence 7, Applicatio
45	78	3.5	1612	2	PCT-US94-0 Sequence 48, Applicati

Sequence 1, Application US/08750532  
Patent No. 5756339  
GENERAL INFORMATION:  
APPLICANT: MITTA, Masanori  
APPLICANT: YAMAMOTO, Katsuhiko  
APPLICANT: MORISHITA, Mio  
APPLICANT: ASADA, Kiyozo  
APPLICANT: TSUNASAWA, Susumu  
APPLICANT: KATO, Ikumoshia  
TITLE OF INVENTION: HYPERHERMOSTABLE PROTEASE GENE  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: BROWDY AND NEIMARK, P. L.L.C.  
STREET: 419 Seventh Street N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States of America  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/750,532  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP95/01095  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 1994/130236  
FILING DATE: 13-JUN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 1994/173912  
FILING DATE: 26-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.

CC	TELEPHONE: (202) 628-5197
CC	TELEFAX: (202) 737-3528
CC	INFORMATION FOR SEQ ID NO: 9:
CC	SEQUENCE CHARACTERISTICS:
CC	LENGTH: 1398 amino acids
CC	TYPE: amino acid
CC	STRANDEDNESS: single
CC	TOPOLOGY: linear
CC	MOLECULE TYPE: peptide
SQ	SEQUENCE 1398 AA; 154545 MW; 10958156 CN;
Dd	Query Match 4.4%; Score 97; DB 1; Length 1398; Best Local Similarity 29.3%; Pred. No. 6.92e+00;
Oy	Matches 17; Conservative 15; Mismatches 22; Indels 4; Gaps 4  743 FVLARKYVEGLEPGLVGRRIIDPTTPVIEDEILNTIYIPEKFTPENNTLTWYDI 800      ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: 745 FVLARVAQ-DP-PELVFKVDVTGAYADT-IPQRL-TLVIASIIKPQNTYCVARRAV 768
RESULT 3	
ID US-08-526-964-4	STANDARD; PRT; 785 AA.
AC xxxxxx	
XX	
DT	
Sequence 4, Application US/08526964	
DE	
CC Sequence 4, Application US/08526964	
CC Patent No. 5698421	
CC GENERAL INFORMATION:	
CC APPLICANT: Lambowitz, Alan M	
CC APPLICANT: Zimmerly, Steven	
CC APPLICANT: Guo, Huatao	
CC APPLICANT: Yang, Jian	
CC TITLE OF INVENTION: Nucleotide Integrase Preparation	
CC NUMBER OF SEQUENCES: 17	
CC CORRESPONDENCE ADDRESS:	
CC ADDRESSEE: Calfee, Halter & Griswold	
CC STREET: 800 Superior Avenue	
CC City: Cleveland	
CC STATE: Ohio	
CC COUNTRY: USA	
CC ZIP: 44114	
CC COMPUTER READABLE FORM:	
CC MEDIUM TYPE: Floppy disk	
CC COMPUTER: IBM PC compatible	
CC OPERATING SYSTEM: PC-DOS/MS-DOS	
CC SOFTWARE: Patentin Release #1.0, Version #1.30	
CC CURRENT APPLICATION DATA:	
CC APPLICATION NUMBER: US/08/526,964	
CC FILING DATE:	
CC CLASSIFICATION: 435	
CC ATTORNEY/AGENT INFORMATION:	
CC NAME: Golrick, Mary E	
CC REGISTRATION NUMBER: 34,829	
CC REFERENCE/DOCKET NUMBER: 22727/00127	
CC TELECOMMUNICATION INFORMATION:	
CC TELEPHONE: (216) 622-8200	
CC TELEFAX: (216) 241-0816	
CC INFORMATION FOR SEQ ID NO: 4:	
CC SEQUENCE CHARACTERISTICS:	
CC LENGTH: 785 amino acids	
CC TYPE: amino acid	
CC STRANDEDNESS: single	
CC TOPOLOGY: linear	
CC MOLECULE TYPE: peptide	
CC HYPOTHETICAL: NO	
SQ SEQUENCE 785 AA; 90390 MW; 3218308 CN;	
Query Match 4.3%; Score 96; DB 1; Length 785; Best Local Similarity 44.4%; Pred. No. 8.14e+00;	



CC APPLICANT: McCaffrey, Patricia  
CC APPLICANT: Jain, Jugnu  
CC TITLE OF INVENTION: NF-ATP, A T LYMPHOCYTE  
CC TITLE OF INVENTION: DNA-BINDING PROTEIN  
CC NUMBER OF SEQUENCES: 21  
CC CORRESPONDENCE ADDRESSES:  
CC ADDRESSEE: Fish & Richardson P.C.  
CC STREET: 225 Franklin Street  
CC CITY: Boston  
CC STATE: Massachusetts  
CC COUNTRY: U.S.A.  
CC ZIP: 02110-2804  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
CC COMPUTER: IBM PS/2 Model 502 or 555X  
CC OPERATING SYSTEM: MS-DOS (Version 5.0)  
CC SOFTWARE: WordPerfect (Version 5.1)  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/145,006C  
CC FILING DATE: October 29, 1993  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/017,052  
CC FILING DATE: February 11, 1993  
CC APPLICATION NUMBER: 08/006,067  
CC FILING DATE: January 15, 1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Fraser, Janis K.  
CC REGISTRATION NUMBER: 34,819  
CC REFERENCE/DOCKET NUMBER: 04590/007001  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 542-5070  
CC TELEFAX: (617) 542-8906  
CC TELEX: 200154  
CC INFORMATION FOR SEQ ID NO: 5:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 890  
CC TYPE: amino acid  
CC STRANDEDNESS:  
CC TOPOLOGY: linear  
CC SEQUENCE 890 AA; 96929 MW; 4192360 CN;  
SQ  
Query Match 3.8%; Score 85; DB 1; Length 890;  
Best Local Similarity 38.8%; Pred. No. 4.65e+01;  
Matches 19; Conservative 13; Mismatches 11; Indels 6; Gaps 4;  
DB 568 AR-YQQSSPAALYQKSKSLPGLIGTQQ--PSLAAP-LGLADAHRS 611  
QY 664 ARFRRERKRAERLRSVKALFS-VLNYERARRPGLIGASVGLDDIHR 711  
RESULT 7  
ID US-08-202-389-23 STANDARD; PRT; 95 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE Sequence 23, Application US/08202389  
XX  
CC Sequence 23, Application US/08202389  
CC Patent No. 5536636  
CC GENERAL INFORMATION:  
CC APPLICANT: Freeman Jr., Robert M.  
CC APPLICANT: Plutzky, Jorge  
CC APPLICANT: Neel, Benjamin G.  
CC APPLICANT: Rosenberg, Robert D.  
CC TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE  
CC TITLE OF INVENTION: PHOSPHATASES HAVING SH2 DOMAINS  
CC NUMBER OF SEQUENCES: 54  
CC CORRESPONDENCE ADDRESSES:  
CC ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
CC STREET: Two Militia Drive

CC CITY: Lexington  
CC STATE: MA  
CC COUNTRY: USA  
CC ZIP: 02173  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/202,389  
CC FILING DATE: 28-FEB-1994  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/983,926  
CC FILING DATE: 01-DEC-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/829,141  
CC FILING DATE: 31-JAN-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/721,112  
CC FILING DATE: 26-JUN-1991  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Granahan, Patricia  
CC REGISTRATION NUMBER: 32,227  
CC REFERENCE/DOCKET NUMBER: BIH92-05WA  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 861-6240  
CC TELEFAX: (617) 861-9540  
CC INFORMATION FOR SEQ ID NO: 23:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 95 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 95 AA; 11151 MW; 48023 CN;  
SQ  
Query Match 3.7%; Score 82; DB 1; Length 95;  
Best Local Similarity 22.9%; Pred. No. 7.38e+01;  
Matches 16; Conservative 23; Mismatches 26; Indels 3; Gaps 3;  
DB 16 YVY-VPDIAEYN-MPQYKREKFVTDARDSSRTVRFQFIDMPEOG-VPKSGEGFIDET 72  
QY 729 YEVKVVOTGAVDIDIPDRDLLEVIAISLIKQNTVCVRAVVOKAANGHKKAFKSHVSTL 788  
DB 73 GQYHKTKEQF 82  
QY 789 TDLQPYMRQF 798  
RESULT 8  
ID PCT-US93-07213-2 STANDARD; PRT; 858 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE Sequence 2, Application PC/TUS9307213  
XX  
CC Sequence 2, Application PC/TUS9307213  
CC GENERAL INFORMATION:  
CC APPLICANT: The Government of the United States of  
CC APPLICANT: America, as represented by The Secretary  
CC TITLE OF INVENTION: NUCLEIC ACIDS AND AMINO ACID  
CC TITLE OF INVENTION: SEQUENCES FOR A STRONGLY IMMUNOREACTIVE PROTEIN ENCODED  
CC NUMBER OF SEQUENCES: 16  
CC CORRESPONDENCE ADDRESSES:  
CC ADDRESSEE: The Government of the United States of  
CC ADDRESSEE: America, as represented by The Secretary  
CC STREET: 6011 Executive Blvd., Suite 325  
CC CITY: Rockville  
CC STATE: Maryland

```

CC      REFERENCE/DOCKET NUMBER: B0801/7020
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: 617-720-3500
CC      TELEFAX: 617-720-2441
CC      INFORMATION FOR SEQ ID NO: 2:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 1178 amino acids
CC      TYPE: amino acid
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: protein
CC      SEQUENCE 1178 AA; 129714 MW; 7185561 CN;
SQ

Query Match      3.7%; Score 81; DB 1; Length 1178;
Best Local Similarity 27.1%; Pred. No. 8.59e+01;
Matches 13; Conservative 14; Mismatches 18; Indels 3; Gaps 3

Db      631 RIRASTVAPGLGYFG-MSMAGFP-IGSDGADITVGLGAVVFRSR 676
      1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11 1:11
Oy      718 RVRADDPPEL-YEVKVDVTGATITIPDRRLTEVIASIKQNTYCVR 764

RESULT 10
ID      PCT-US95-02044-2      STANDARD;      PRT;      1178 AA.
XX      xxxxxx
AC
DE      Sequence 2, Application PC/TUS9502044
XX      DE
DT      Sequence 2, Application PC/TUS9502044
CC      GENERAL INFORMATION:
CC      APPLICANT:
CC      TITLE OF INVENTION: Novel Integrin alpha subunit
CC      NUMBER OF SEQUENCES: 25
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: Wolf, Greenfield and Sacks, P.C.
CC      STREET: 600 Atlantic Avenue
CC      CITY: Boston
CC      STATE: MA
CC      COUNTRY: USA
CC      ZIP: 02210
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Floppy disk
CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: Patent In Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: PCT/US95/02044
CC      FILING DATE: herewith
CC      CLASSIFICATION:
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: 08/199,776
CC      FILING DATE: 18 February 1994
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: Plumer, Elizabeth R.
CC      REGISTRATION NUMBER: 36,637
CC      REFERENCE/DOCKET NUMBER: B0801/7020
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: 617-720-3500
CC      TELEFAX: 617-720-2441
CC      INFORMATION FOR SEQ ID NO: 2:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 1178 amino acids
CC      TYPE: amino acid
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: protein
CC      SEQUENCE 1178 AA; 129714 MW; 7185561 CN;
SQ

Query Match      3.7%; Score 81; DB 2; Length 1178;
Best Local Similarity 27.1%; Pred. No. 8.59e+01;
Matches 13; Conservative 14; Mismatches 18; Indels 3; Gaps 3

```



CC APPLICANT: Beck, James Joseph  
CC APPLICANT: Hill, Dwight Steven  
CC APPLICANT: Ryals, John Andrew  
CC APPLICANT: Gaffney, Thomas Deane  
CC APPLICANT: Lam, Stephen Ting  
CC APPLICANT: Hammer, Phillip E.  
CC APPLICANT: Uknes, Scott Joseph  
CC TITLE OF INVENTION: Genes for the synthesis of  
CC antipathogenic substances  
CC NUMBER OF SEQUENCES: 22  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Ciba-Geigy Corporation  
CC STREET: 7 Skyline Drive  
CC CITY: Hawthorne  
CC STATE: NY  
CC COUNTRY: USA  
CC ZIP: 10532  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/458,076A  
CC FILING DATE: 01-JUN-1995  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/457,205  
CC FILING DATE: 01-JUN-1995  
CC APPLICATION NUMBER: 08/258,261  
CC FILING DATE: 08-JUN-1994  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Elmer, James Scott  
CC REGISTRATION NUMBER: 36,129  
CC REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 919-541-8614  
CC TELEFAX: 919-541-8689  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 567 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 567 AA; 65037 MW; 1552397 CN;  
SQ  
Query Match 3.6%; Score 79; DB 1; Length 567;  
Best Local Similarity 25.0%; Pred. No. 1.16e+02;  
Matches 8; Conservative 13; Mismatches 11; Indels 0; Gaps 0;  
DB 141 LLAQAIRYCKVHOKTTEYHADKGVAVTT 172  
Y 909 LVKGVPEYGCVVNLKRTVNFPEDEALGSTA 940  
RESULT 14  
ID US-08-258-261B-4 STANDARD; PRT; 567 AA.  
XX xxxxxx  
DT  
XX  
XX Sequence 4, Application US/08258261B  
DE Sequence 4, Application US/08258261B  
CC Patent No. 563949  
CC GENERAL INFORMATION:  
CC APPLICANT: Schnupp, Thomas  
CC APPLICANT: Ligon, James M.  
CC APPLICANT: Beck, James Joseph  
CC APPLICANT: Hill, Dwight Steven  
CC APPLICANT: Ryals, John Andrew  
CC APPLICANT: Gaffney, Thomas Deane

CC APPLICANT: Lam, Stephen Ting  
CC APPLICANT: Hammer, Phillip E.  
CC APPLICANT: Uknes, Scott Joseph  
CC TITLE OF INVENTION: Genes for the synthesis of  
CC antipathogenic substances  
CC NUMBER OF SEQUENCES: 22  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Ciba-Geigy Corporation  
CC STREET: 7 Skyline Drive  
CC CITY: Hawthorne  
CC STATE: NY  
CC COUNTRY: USA  
CC ZIP: 10532  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/258,261B  
CC FILING DATE: 08-JUN-1994  
CC CLASSIFICATION: 800  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/457,205  
CC FILING DATE: 01-JUN-1995  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Elmer, James Scott  
CC REGISTRATION NUMBER: 36,129  
CC REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 919-541-8614  
CC TELEFAX: 919-541-8689  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 567 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 567 AA; 65037 MW; 1552397 CN;  
SQ  
Query Match 3.6%; Score 79; DB 1; Length 567;  
Best Local Similarity 25.0%; Pred. No. 1.16e+02;  
Matches 8; Conservative 13; Mismatches 11; Indels 0; Gaps 0;  
DB 141 LLAQAIRYCKVHOKTTEYHADKGVAVTT 172  
Y 909 LVKGVPEYGCVVNLKRTVNFPEDEALGSTA 940  
RESULT 15  
ID US-08-457-342-4 STANDARD; PRT; 567 AA.  
XX xxxxxx  
DT  
XX  
XX Sequence 4, Application US/08457342  
DE Sequence 4, Application US/08457342  
CC Patent No. 5662898  
CC GENERAL INFORMATION:  
CC APPLICANT: Schnupp, Thomas  
CC APPLICANT: Ligon, James M.  
CC APPLICANT: Beck, James Joseph  
CC APPLICANT: Hill, Dwight Steven  
CC APPLICANT: Ryals, John Andrew  
CC APPLICANT: Gaffney, Thomas Deane  
CC APPLICANT: Lam, Stephen Ting  
CC APPLICANT: Hammer, Phillip E.  
CC APPLICANT: Uknes, Scott Joseph  
CC TITLE OF INVENTION: Genes for the synthesis of  
CC antipathogenic substances  
CC NUMBER OF SEQUENCES: 22





QY 239 PAPGARRRGGSASRS-LPLPKRR-RCGAAPERTPVGGSGMAHGRTRGSPSD-RGFCV 295  
 Db 96 tppappedgrrpggaagaastrdgrrsggrrpprpskappxerw 140  
 QY 296 SPAPPAEATSLGALSGTRHSHPSVGRQHHAGPPSTRP-PRW 339

## RESULT

ID W18664 standard; Protein; 386 AA.  
 AC W18664;  
 DT 24-JUL-1997 (first entry)  
 DE Fragmented human NF-H gene +1 frameshift mutant product.  
 KW Frameshift mutation product; GAGA motif; somatic mutation; diagnosis;  
 KW detection; antibody; probe; cancer; neoplasia; neurodegenerative;  
 KW Parkinson's; Alzheimer's disease; Pick's; Huntington's disease;  
 KW Down's syndrome; frontal lobe dementia; progressive supranuclear palsy;  
 KW PSP; amyotrophic lateral sclerosis; multiple sclerosis; MS;  
 KW cardiovascular; rheumatoid arthritis.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT misc\_difference 1..387  
 FT /note="X corresponds to a stop codon in the  
 FT accompanying DNA file, T69796."  
 FT 129..135  
 FT /note="antigenic peptide used for antibody  
 FT production"

MO9712993-A2.

PD 10-APR-1997.  
 PE 02-OCT-1996; IB1106.  
 PR 02-OCT-1995; GB-020080.  
 PR 11-JAN-1996; US-009832.  
 PA (ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.  
 PA (UYUR-) UNIV ROTTERDAM ERASMUS.  
 PI Burbach JPH, Grosveld FG, Van Leeuwen FW;  
 DR N-PADB; T69796.  
 DR N-PADB; T69796.  
 PT Use of mutant genes having frame:shift mutation(s) - for developing  
 PT prods. for the diagnosis, prevention and treatment of associated  
 PT diseases, e.g. cancer or neurodegenerative disease  
 PS Claim 22: Fig 9; 123pp; English.  
 CC W18663 and W18664 are +2 and +1 frameshift mutations, respectively,  
 CC of a sequence comprising fragments of the coding sequence of the  
 CC human neurofilament subunit NF-H gene corresponding to nucleotides  
 CC 1-1162 of the wild-type NF-H gene. This region contains GAGAG motifs.  
 CC Frameshift mutants of the tau, ubiquitin, apolipoprotein E,  
 CC microtubule-associated protein 2 (MAP-2), neurofilament subunit L, M  
 CC and H and amyloid A4 genes are claimed. All these genes share a common  
 CC GAGAN motif (N-A, G, C or T), which is the site of common GA  
 CC dinucleotide deletion(s) that cause neurodegenerative disorders.  
 CC Antigenic peptides used for the production of antibodies, and small  
 CC nucleic acid sequences derived from frameshift mutants are used in the  
 CC diagnosis, prevention and treatment of cancer and neurodegenerative  
 CC diseases, e.g. Parkinson's disease, Alzheimer's disease, Down's  
 CC syndrome, frontal lobe dementia (Pick's disease), progressive  
 CC supranuclear palsy (PSP), amyotrophic lateral sclerosis, Huntington's  
 CC disease, multiple sclerosis, and other degenerative diseases such as  
 CC cardiovascular disease and rheumatoid arthritis.  
 SQ Sequence 386 AA;

Query Match 1.4%; Score 117; DB 22; Length 386;  
 Best Local Similarity 28.0%; Pred. No. 5.36e+00;  
 Matches 30; Conservative 26; Mismatches 50; Indels 1; Gaps 1;

Db 228 aggplry-gravrtarprdrtrgaaprgarsatpgagaparrtaraprrrrpaary 286  
 QY 232 AGVPLGAPAGARRRGGSASRSLPLPKRRGAAPERTPVGGSGMAHGRTRGSPSDRG 291  
 Db 287 grgggpragallraggrgrrgaagagagavrlpaapprrgrra 333  
 QY 292 FCVVSPPAPPAEATSLGALSGTRHSHPSVGRQHHAGPPSTRP 338

## RESULT

ID R23846 standard; Protein; 399 AA.  
 AC R23846;  
 DT 05-NOV-1992 (first entry)  
 DE Prepro form of human osteogenic polypeptide-2.  
 KW Human osteogenic polypeptide-2; prepro sequence; implant;  
 KW endochondral bone formation; skeletal abnormality; dental abnormality;  
 KW non-union fracture; cartilage repair; osteoarthritis;  
 KW periodontal applicn.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT cleavage\_site 13..18  
 FT /label= cleavage site for signal peptide removal  
 FT 257..261  
 FT cleavage\_site  
 FT /label= cleavage site to form mature protein  
 FT 261..399  
 FT /label= mature protein 'hOP2-A'  
 FT 264..399  
 FT /label= truncated protein 'hOP2-P'  
 FT 267..399  
 FT /label= truncated protein 'hOP2-R'  
 FT /label= truncated protein 'hOP2-S'  
 FT /note="J"  
 FT 240..399  
 FT /label= conserved six cysteine skeleton  
 FT 297..399  
 FT /label= conserved seven cysteine skeleton

MO9207073-A.

PD 30-APR-1992.  
 PE 18-OCT-1991; U07635.  
 PR 18-OCT-1990; US-599543.  
 PA (CREA-) CREATIVE BIOMOLECULES INC.  
 PI Kuberassampath T, Oppermann H, Ozkaynak E, Rueger DC;  
 DR N-PADB; N24522.  
 DR N-PADB; N24522.  
 PT New osteogenic polypeptide(s) and their dimeric proteins - useful  
 PT in association with a matrix for bone reconstitution after  
 PT orthopaedic and plastic surgery  
 PS Disclosure: 51-53; 69pp; English.  
 CC The sequence encodes the prepro form of human osteogenic polypeptide-2  
 CC (hOP2-PP). The protein is useful as a subunit of osteogenic  
 CC proteins capable of endochondral bone formation, and allogenic and  
 CC xenogenic implants in mammals. When implanted with a matrix, the  
 CC polypeptide locally induces the full developmental cascade of  
 CC endochondral bone formation and bone marrow differentiation. This  
 CC can be used for optimal predictable bone formation to correct, eg  
 CC acquired or congenital craniofacial and other skeletal or dental  
 CC anomalies, and to treat non-union fractures. The polypeptide also  
 CC has dental and periodontal applications, and may be used for cartilage  
 CC repair, eg in osteoarthritis treatment. See also WO9105802.  
 SQ Sequence 399 AA;

Query Match 1.4%; Score 118; DB 1; Length 399;  
 Best Local Similarity 39.7%; Pred. No. 4.60e+00;  
 Matches 25; Conservative 13; Mismatches 19; Indels 6; Gaps 6;

Db 12 glalcalggggpplppp-gc-pqrrlg-arddvq-rellavlgjpr-prrppaas 66  
 QY 194 GPPLYQLGATQA-RPPHNSGPRRRRLGRCERAMNHSREAGVGLGAPGARRRGGSASR 252  
 Db 67 rlp 69  
 QY 253 SLP 255

## RESULT

ID W40193 standard; Protein; 399 AA.  
 AC W40193;  
 DT 08-JUN-1998 (first entry)  
 DE Mouse OP-2 protein.  
 KW Osteogenic protein; OP-2; bone morphogenic protein; OP/BMP family;  
 KW chronic renal failure; renal therapeutic agent; disease; diabetes;

KW nephropathy; glomerulopathy; hypertrophy; sclerosis; nephritis; mouse;  
 KW dysplasia; fibrosis; glomerular filtration rate; GFR.  
 OS Mus sp.  
 FT Key Location/Qualifiers  
 FT Peptide 1..17 /label= signal  
 FT Peptide 18..263 /label= propeptide  
 FT Protein 264..402 /label= Op-2  
 PN W09741881-A1.  
 PD 13-NOV-1997.  
 PE 06-MAY-1997; U07816.  
 PR 06-MAY-1996; US-643321.  
 PA (CREA-) CREATIVE BIOMOLECULES INC.  
 PI Cohen CM, Sampath KT;  
 DR WPI: 97-558690/51.  
 DR N-PSDB: V10348.  
 PT Treatment of chronic renal failure - using an osteogenic  
 PT protein/bone morphogenetic protein renal therapeutic agent or  
 PT morphogen or renal mesenchymal progenitor cells  
 PS Disclosure: Page 68-69; 113pp: English.  
 CC This sequence represents the mouse osteogenic protein, OP-2. This  
 CC protein is used in a method for the treatment of a mammal having, or at  
 CC risk of, chronic renal failure which comprises administering an  
 CC osteogenic protein/bone morphogenetic protein (OP/BMP) renal therapeutic  
 CC agent or morphogen. The method can be used for treating e.g. chronic  
 CC renal failure, end-stage renal disease, chronic diabetic nephropathy,  
 CC diabetic glomerulopathy, diabetic renal hypertrophy, hypertensive  
 CC nephrosclerosis, hypertensive glomerulosclerosis, chronic  
 CC glomerulonephritis, hereditary nephritis, renal dysplasia, or a patient  
 CC afflicted with e.g. glomerular hypertrophy, tubular hypertrophy,  
 CC glomerulosclerosis, tubulointerstitial sclerosis or renal fibrosis. Such  
 CC therapeutic agents can prevent, inhibit or delay the progressive loss of  
 CC functional nephron units and the progressive decline in glomerular  
 CC filtration rate (GFR) which slowly but inevitably leads to the need for  
 CC renal replacement therapy.  
 SQ Sequence 399 AA;

Query Match 1.4%; Score 118; DB 29; Length 399;  
 Best Local Similarity 38.7%; Pred. No. 4.60e+00;  
 Matches 24; Conservative 12; Mismatches 22; Indels 4; Gaps 4;

Db 12 glalcalgsgghpprp-htc-pqrrlg-arerrmdgrelavlgprr-prpragpaar 67  
 QY 194 GPPLTQLGAATQARPPHAGSPRRRLGGERAMNHSVREAGVPLGAPAGARRGGSASRS 253

Db 68 qp 69  
 QY 254 LP 255

RESULT 5  
 ID W44314 standard; Protein: 399 AA.  
 AC W44314:  
 DT 27-MAY-1998 (first entry)  
 DE Mouse osteogenic protein OP2.  
 KW Mouse; osteogenic protein; subunit; endochondral bone formation;  
 KW dimeric; recombinant protein.  
 OS Mus sp.  
 PN US5714589-A.  
 PD 03-FEB-1998.  
 PE 23-MAY-1995; 447570.  
 PR 21-FEB-1992; US-841646.  
 PR 08-APR-1988; US-179406.  
 PR 15-AUG-1988; US-232630.  
 PR 23-FEB-1989; US-315342.  
 PR 17-OCT-1989; US-422613.  
 PR 17-OCT-1989; US-422699.  
 PR 22-FEB-1990; US-483913.  
 PR 20-AUG-1990; US-569920.  
 PR 07-SEP-1990; US-579665.  
 PR 18-OCT-1990; US-599543.

PR 18-OCT-1990; US-600024.  
 PR 21-NOV-1990; US-616374.  
 PR 04-DEC-1990; US-621849.  
 PR 04-DEC-1990; US-621988.  
 PR 22-FEB-1991; US-660162.  
 PR 20-DEC-1991; US-810560.  
 PR 28-JAN-1992; US-827052.  
 PR 01-NOV-1993; US-147023.  
 PR 23-MAY-1995; US-447570.  
 PA (STYC ) STRYKER CORP.  
 PI Kuderampsampath T, Oppermann H, Ozkaynak E, Pang RHL,  
 PI Rueger DC;  
 DR WPI: 98-158353/14.  
 DR N-PSDB: V15217.  
 PT Extraction of osteogenic protein from mixture - using antibodies  
 PT specific for novel polypeptide chains useful as subunit(s) of  
 PT dimeric osteogenic protein(s)  
 PS Disclosure: Column 137-140; 127pp: English.  
 CC The present sequence represents a mouse osteogenic protein, which is  
 CC used in the present invention. The present invention describes methods  
 CC for selectively extracting an osteogenic protein (OP) from a mixture.  
 CC The method comprises: (a) exposing the mixture to an antibody that  
 CC specifically binds OP, separating the resulting antibody-protein complex  
 CC from the mixture, and dissociating the complex. In the methods OP  
 CC comprises a pair of oxidised subunits that are disulphide-bonded to form  
 CC a dimer, and one of the subunits has an amino acid (aa) sequence  
 CC sufficiently homologous to residues 335-431 of a 431 aa protein  
 CC designated OP5, sequence given in the specification. In dimeric form OP  
 CC is capable of inducing cartilage and endochondral bone formation in a  
 CC mammal when disposed within a matrix implanted in the mammal. The  
 CC methods are used for recovering the recombinant proteins from cell  
 CC cultures.  
 SQ Sequence 399 AA;

Query Match 1.4%; Score 118; DB 29; Length 399;  
 Best Local Similarity 38.7%; Pred. No. 4.60e+00;  
 Matches 24; Conservative 12; Mismatches 22; Indels 4; Gaps 4;

Db 12 glalcalgsgghpprp-htc-pqrrlg-arerrmdgrelavlgprr-prpragpaar 67  
 QY 194 GPPLTQLGAATQARPPHAGSPRRRLGGERAMNHSVREAGVPLGAPAGARRGGSASRS 253

Db 68 qp 69  
 QY 254 LP 255

RESULT 6  
 ID R85759 standard; Protein: 399 AA.  
 AC R85759:  
 DT 19-JUN-1996 (first entry)  
 DE MOP-2.  
 KW Human; osteogenic protein; hOP-1; murine; MOP-1; TGF-beta superfamily;  
 KW transforming growth factor-beta; dimer; antibody; epitope; hippocampus;  
 KW purification; implantable osteogenic device; bone formation; craniofacial;  
 KW anomaly; skeletal; dental; endochondral bone formation;  
 KW non-union fracture; cartilage repair; osteoarthritis.  
 OS Mus musculus.  
 PN US5468845-A.  
 PD 21-NOV-1995.  
 PE 08-APR-1988; US-179406.  
 PR 15-AUG-1988; US-232630.

FT Key Location/Qualifiers  
 FT Peptide 1..260 /note= "Prepro-peptide"  
 FT Protein 261..399 /note= "Mature MOP-2, also designated MOP2-ALA,  
 FT Peptide 303..399 claim 5"  
 FT Peptide 17..260 /note= "MOP-2 peptide, claim 5"  
 FT Peptide 17..260 /note= "MOP-2 peptide, claim 5"

23-FEB-1989; US-315342.  
 PR 17-OCT-1989; US-422699.  
 PR 17-OCT-1989; US-422613.  
 PR 22-FEB-1990; US-483913.  
 PR 20-AUG-1990; US-569920.  
 PR 07-SEP-1990; US-579865.  
 PR 18-OCT-1990; US-599543.  
 PR 21-NOV-1990; US-616374.  
 PR 04-DEC-1990; US-621849.  
 PR 04-DEC-1990; US-621988.  
 PR 22-FEB-1991; US-660162.  
 PR 20-DEC-1991; US-810560.  
 PR 28-JAN-1992; US-827052.  
 PR 21-FEB-1992; US-841646.  
 PR 01-NOV-1993; US-147023.  
 PA (STY) STRIKER CORP  
 PI Kuberampath T, Oppermann H, Ozkaynak E, Rueger DC;  
 DR N-PSDB: T02599.  
 DR N-PSDB: T02599.  
 PT Antibodies with osteogenic protein binding specificity - used in  
 PT purification of osteogenic proteins, and as antigenic proteins  
 PS Disclosure; Column 139-42; 129pp; English.  
 CC This sequence represents the murine osteogenic protein, mop-2. mop-2  
 CC has homology with proteins in the TGF-beta superfamily. Fragments  
 CC of the mop-2 protein can be used in the production of dimeric  
 CC peptides which may be used in the generation of antibodies with  
 CC binding specificities for osteogenic proteins. The antibodies are  
 CC capable of binding specifically to an epitope of the osteogenic protein  
 CC and may be used in purification protocols. Osteogenic proteins, such as  
 CC mop-2, may be used in an implantable osteogenic device which allows  
 CC predictable bone formation to correct acquired and congenital  
 CC craniofacial and other skeletal or dental anomalies. They may be used to  
 CC induce local endochondral bone formation in non-union fractures and in  
 CC other clinical applications including dental and periodontal applications  
 CC where bone formation is required. Other potential applications include  
 CC cartilage repair, e.g. in the treatment of osteoarthritis.  
 CC Sequence 399 AA;  
 SQ

Query Match 1.4%; Score 118; DB 16; Length 399;  
 Best Local Similarity 38.7%; Pred. No. 4.60e+00;  
 Matches 24; Conservative 12; Mismatches 22; Indels 4; Gaps 4;

Db 12 glalcalggghgppr-htc-pqrrlg-arerrdmqrellavlgpgr-prpragpaar 67  
 | : | | : | | | | : | | | | | : | | | | : | | : | | : | |  
 QY 194 GPPLYQLGAAIQARPPPHASGPRRLCCERAMNHSVRAGVPLGLPARGARRRGSASRS 253

Db 68 qp 69  
 |  
 QY 254 LP 255

RESULT 7  
 ID R54938 standard; Protein; 399 AA.  
 AC R54938;  
 DT 15-OCT-1994 (first entry)  
 DE Osteogenic protein mop2-pp.  
 KW Morphogenic protein; mop-2-pp; OP-2; mop2; mop-2;  
 KW tissue morphogenesis; osteogenic protein.  
 OS Mus sp.  
 PN MO9410203-A.  
 PD 11-MAY-1994.  
 PE 02-NOV-1993; U10520.  
 PR 03-NOV-1992; US-971091.  
 PR 04-MAR-1993; US-029335.  
 PR 31-MAR-1993; US-040510.  
 PA (CREA-) CREATIVE BIOMOLECULES INC.  
 PI Cohen CM, Kuberampath T, Oppermann H, Ozkaynak E;  
 PI Pang RHL, Rueger DC;  
 DR WPI: 94-167392/20.  
 DR P-PSDB: 065394.  
 PT A morphogenically active protein mop-3 - for inducing tissue  
 PT morphogenesis in mammals  
 PS Disclosure; Page 134-136; 164pp; English.

CC A novel mouse morphogenic protein, OP3, has the sequence given in  
 CC R54934, and is encoded by cDNA of sequence 065390. cDNA and protein  
 CC sequences were also provided for human osteogenic protein OP1  
 CC (065391, R54935), mouse OP1 (065392, R54936), human OP2 (065393,  
 CC R54937) and mouse OP2 (065394, R54938), as well as the genomic DNA  
 CC sequence of human OP2 (065395). Generic sequences given in R54939-  
 CC 40 accommodate homologies between OP1, OP2, OP3 and other morphogen  
 CC protein family members.  
 CC Sequence 399 AA;  
 SQ

Query Match 1.4%; Score 118; DB 10; Length 399;  
 Best Local Similarity 38.7%; Pred. No. 4.60e+00;  
 Matches 24; Conservative 12; Mismatches 22; Indels 4; Gaps 4;

Db 12 glalcalggghgppr-htc-pqrrlg-arerrdmqrellavlgpgr-prpragpaar 67  
 | : | | : | | | | : | | | | | : | | | | : | | : | | : | |  
 QY 194 GPPLYQLGAAIQARPPPHASGPRRLCCERAMNHSVRAGVPLGLPARGARRRGSASRS 253

Db 68 qp 69  
 |  
 QY 254 LP 255

RESULT 8  
 ID R33908 standard; Protein; 399 AA.  
 AC R33908;  
 DT 21-JUL-1993 (first entry)  
 DE Mouse osteogenic protein 2 (mop-2).  
 KW Bone; loss; increase; fracture; post-menopausal; senile;  
 KW osteoporosis; hyperparathyroidism; skeletal microstructure defects;  
 KW chronic renal failure; kidney disease; osteomalacia; vitamin D;  
 KW deficiency-induced osteopenia, osteoporosis; Paget's disease;  
 KW bone mass; imbalance; resorption; formation; dialysis; calcium;  
 KW phosphate; metabolism; murine.  
 OS Mus musculus.  
 FH Key  
 FT protein  
 FT 261..399  
 FT Location/Qualifiers  
 FT region  
 FT 18..260  
 FT /note="mature protein"  
 FT /note="pro region, cleaved to yield mature,  
 FT active protein"  
 FT 298..299  
 FT /note="conserved seven cysteine skeleton"

region  
 FT MO9305751-A.  
 PD 01-APR-1993.  
 PE 28-AUG-1992; U07432.  
 PR 30-AUG-1991; US-752764.  
 PR 30-AUG-1991; US-752857.  
 PR 31-JUL-1992; US-923780.  
 PA (CREA-) CREATIVE BIOMOLECULES INC.  
 PI Cohen CM, Kuberampath T, Oppermann H, Ozkaynak E,  
 PI Pang RHL, Rueger DC, Smart JE;  
 DR WPI: 93-117208/14.  
 PT Use of morphogenic or in-vivo morphogenic-stimulating agent - to  
 PT prevent bone loss or increase, used for treating bone fractures,  
 PT post-menopausal or senile osteoporosis, hyperparathyroidism etc.  
 PS Disclosure; Page 122-123; 162pp; English.  
 CC The sequence is that of mouse osteogenic protein 2 (mop-2) a  
 CC morphogenically active protein which may be used as part of a  
 CC method for treating a bone fracture or a disease which causes or  
 CC results in bone fractures or other defects in skeletal  
 CC microstructure. Such diseases include chronic renal failure and  
 CC other kidney diseases, osteomalacia, vitamin D deficiency-induced  
 CC osteopenia or osteoporosis, postmenopausal or senile osteoporosis,  
 CC hyperparathyroidism and Paget's disease. The methods can be used for  
 CC protecting individuals at risk for loss of bone mass such as  
 CC postmenopausal females, aged individuals and individuals undergoing  
 CC dialysis. The loss of bone mass may result from an imbalance in bone  
 CC resorption or bone formation, an imbalance of calcium or phosphate  
 CC metabolism, a vitamin D imbalance or be nutritionally or hormonally  
 CC induced.  
 CC Sequence 399 AA;  
 SQ

Query Match	1.4%;	Score 118;	DB 9;	Length 399;
Best Local Similarity	38.7%;	Pred. No. 4.60e+00;		
Matches	24;	Conservative	12;	Mismatches 22; Indels 4; Gaps 4
Db	12	gla aal gg hg pp p htc p rr lg a er rd mq re lav l p gr p pr eq p aa ar	67	
Oy	194	GPPLTOLGAAATGAARPPPHASGPRRLTGCERAMNHSVREAGVPLGLPAPGARRGGSASRS	253	
Db	68	qp	69	
Oy	254	LP	255	
RESULT	9			
ID	R46744	standard; Protein; 399 AA.		
AC	R46744;			
DT	25-AUG-1994	(first entry)		
DE	Mouse osteogenic protein mop2.			
KW	mouse osteogenic protein; mop2; murine; morphogen;			
KW	infant food formulation; tissue morphogenesis; tissue development;			
KW	bone growth; morphogen-enriched nutritional product.			
OS	Muride.			
PN	W09403075-A.			
PD	17-FEB-1994.			
PE	29-JUL-1993;	U07190.		
PR	31-JUL-1992;	US-923780.		
PR	31-JUL-1992;	US-922813.		
PR	16-SEP-1992;	US-946235.		
PR	04-MAR-1993;	US-029335.		
PR	31-MAR-1993;	US-040510.		
PA	(CREA-) CREATIVE BIOMOLECULES INC.			
PI	Jones WK, Ruberasampath T, Oppermann H, Ozkaynak E;			
PI	Rueger DC, Tucker RF, Cohen CM, Pang RHL;			
PI	WPI: 94-065304/08.			
DR	N-PSDB: Q38052.			
PT	Morphogen enriched dietary compositions and infant formula -			
PT	capable of enhancing tissue morphogenesis, development and			
PT	viability, e.g. in infants, aged individuals and metabolic			
PT	disorders, e.g. anorexia nervosa, etc			
PS	Disclosure and claims 25-26; Page 127-129; 160pp; English.			
CC	Murine osteogenic protein mop2 and proteins having at least 70%			
CC	homology with it are preferred morphogens for inclusion in new			
CC	morphogen-enriched nutritional formulations. The formulations			
CC	are dietary compositions suitable for people at risk for tissue			
CC	damage due to protein energy malnutrition or to altered metabolism			
CC	function and infant formulations to enhance tissue development in			
CC	an infant or juvenile.			
CC	Sequence 399 AA;			
CC	Sequence 399 AA;			
Query Match	1.4%;	Score 118;	DB 9;	Length 399;
Best Local Similarity	38.7%;	Pred. No. 4.60e+00;		
Matches	24;	Conservative	12;	Mismatches 22; Indels 4; Gaps 4
Db	12	gla aal gg hg pp p htc p rr lg a er rd mq re lav l p gr p pr eq p aa ar	67	
Oy	194	GPPLTOLGAAATGAARPPPHASGPRRLTGCERAMNHSVREAGVPLGLPAPGARRGGSASRS	253	
Db	68	qp	69	
Oy	254	LP	255	
RESULT	10			
ID	R50201	standard; Protein; 399 AA.		
AC	R50201;			
DT	11-OCT-1994	(first entry)		
DE	Murine OP-2.			
KW	OP-1; OP-2; CBMP2; Vg1(fx); Vgr(fx); DPP(fx);			
KW	GD-1(fx); 60A(fx); BMP3(fx); BMP5(fx); BMP6(fx);			
KW	osteogenic protein; morphogen; morphogenic protein;			
KW	gastrointestinal tract; luminal lining; epithelial cell;			
KW	proliferation; ulcer; lesion; inflammation; regeneration;			

OS	tissue.
KW	Mus musculus.
PN	W09406420-A.
PD	31-MAR-1994.
PF	15-SEP-1993; U08885.
PR	15-SEP-1993; US-945286.
PR	04-MAR-1993; US-029335.
PR	31-MAR-1993; US-040510.
PA	(CREA-) CREATIVE BIOMOLECULES INC.
P1	Charette WF, Cohen CM, Kuderatsepath T, Oppermann H;
P1	Ozkaynak E, Pang RHL, Rueger DC, Smart JE;
DR	N-PSSD: 045119.
PT	Maintaining integrity of gastrointestinal lining using a
PT	morphogen (stimulant) - for treating or preventing ulceration,
PT	also to inhibit endothelial cell proliferation and reduce side
PT	effects of cancer therapy.
PS	Claim 35-36; Page 115-117; 151pp; English.
CC	Morphogens comprising an amino acid sequence sharing at least
CC	70% homology with OP-1, OP-2, BMP2, BMP3(fx), Vgl(fx), Vgr(fx),
CC	BMP(fx), GDF-1(fx), 60A(fx) and at least 80% homology with
CC	DMP(fx) and BMD6(fx) are useful for maintaining the integrity of
CC	the gastrointestinal tract luminal lining in a mammal, including
CC	(1) limiting epithelial cell proliferation, (2) inhibiting ulcerative
CC	lesion formation, (3) inhibiting inflammation normally associated
CC	with ulcerative diseases, and/or (4) stimulating the repair of
CC	ulcerative lesions and the regeneration of the luminal tissue.
SQ	Sequence 399 AA;
Query Match	1.4%; Score 118; DB 10; Length 399;
Best Local Similarity	38.7%; Pred. NO. 4, 60e+00;
Matches 24; Conservative 12; Mismatches 22; Indels 4; Gaps 4;	
Db	12 g1acalcagghqhppp-htc-pgrrl-garetdmqreilaivlgpr-dprpqpaar 67
Oy	194 GPPLYGGAQAQANPAPPASGPERRRLCGERAMNNSVREAGVPGLPAPGARRGGSASRS 253
Db	68 qp 69
Oy	254 LP 255
RESULT 11	
ID	R47253 standard; Protein; 399 AA.
AC	R47253;
DT	15-AUG-1994 (first entry)
DE	MOP2.
KW	Human; hippocampus; osteogenic protein; OP-1; subunit; dimer;
KW	morphogenic activity; cysteine; morphogen; family; pro-region;
KW	complex; soluble; aqueous solvent; therapeutic composition;
KW	syntaxon-allenating; co-factor; antibody; diagnosis; assay;
KW	quantitate; mature.
OS	Mus musculus.
PN	W09403600-A.
PD	17-FEB-1994.
PF	29-JUL-1993; U07189.
PR	31-JUL-1992; US-923780.
PR	04-MAR-1993; US-029335.
PR	31-MAR-1993; US-040510.
PA	(CREA-) CREATIVE BIOMOLECULES INC.
P1	Jones WK, Kuderatsepath T, Oppermann H, Ozkaynak E;
P1	Rueger DC, Tucker RF;
DR	WPI: 94-065689/08.
DR	N-PSSD: 056201.
PT	Morphogenic protein soluble complex - for regeneration of tissue
PT	in mammals and diagnosing tissue disorders
PS	Claim 3; Page 75-77; 120pp; English.
CC	This sequence represents the murine derived protein, osteogenic
CC	protein, mop-2. The mature OP-2 protein was used as at
CC	least one subunit in the dimeric protein of the invention. This
CC	dimeric protein comprises a pair of protein subunits which are
CC	associated to give a structure with morphogenic activity. Each
CC	subunit comprises more than 100 amino acids having a pattern of



RESULT 14  
ID R57974 standard: Protein; 399 AA.  
AC R57974;  
DT 11-OCT-1994 (first entry)  
DE Murine OP-2.  
OP-1: OP-2; CBMP2; Vg1(fx); Vgr(fx); DPP(fx);  
KM GDF-1(fx); 60A(fx); BMP3(fx); BMP5(fx); BMP6(fx);  
KM tooth socket; alveolus; osteogenic protein; morphogen;  
KM morphogenic protein; periodontal tissue; regeneration;  
KM tooth implant; integration; inhibition.  
OS Mus musculus.  
PN WO9406399-A.  
PD 31-MAR-1994.  
PF 15-SEP-1993; U08742.  
PR 15-SEP-1992; US-945285.  
PR 04-MAR-1993; US-029335.  
PR 31-MAR-1993; US-040510.  
PA (CREA-) CREATIVE BIOMOLECULES INC.  
PI Cohen CM, Kuberampath T, Oppermann H, Ozkaynak E;  
PI Pang RHL, Rueger DC, Smart JE;  
DR WPI: 94-118107/14.  
DR N-PSDB; Q67314.  
PT Morphogen-induced periodontal tissue regeneration - used in  
PT integrating as implanted tooth in tooth socket or to inhibit  
PT tissue loss associated with periodontal disease or injury  
PS Claim 28-29: Page 100-103; 132pp; English.  
CC Morphogens comprising an amino acid sequence sharing at least  
CC 70% homology with OP-1, OP-2, CBMP2, Vg1(fx), Vgr(fx), DPP(fx),  
CC GDF-1(fx), 60A(fx) and at least 80% homology with BMP3(fx),  
CC BMP5(fx) and BMP6(fx) are useful for integrating an implanted  
CC tooth in a tooth socket and for inhibiting tissue loss associated  
CC with periodontal disease or injury.  
SQ Sequence 399 AA;

Query Match 1.4%; Score 118; DB 10; Length 399;  
Best Local Similarity 38.7%; Pred. No. 4.60e+00;  
Matches 24; Conservative 12; Mismatches 22; Indels 4; Gaps 4;

Db 12 glalcalagggghppp-htc-pqrrlg-areerdmqrellavlgpgr-prpraqpaar 67  
QY 194 GPPLYQLGAATQARPPPHASGPRRRRLGGERAMNHSVREAGVPLGLPAPGARRRGSASRS 253  
Db 68 qp 69  
QY 254 LP 255

RESULT 15  
ID R33411 standard: Protein; 399 AA.  
AC R33411;  
DT 15-JUL-1993 (first entry)  
DE Murine Pro-OP-2-PP.  
KM morphogenic; osteogenic protein; developmental cascade; mop-2;  
KM mouse; inflammation; anti-inflammatory; Transforming Growth Factor;  
KM TGF-beta super-family; hippocampus.  
OS Mus.  
FH Key Location/Qualifiers  
FT cds 93..1292 /\*tag- a  
FT /\*note- "mop-2-pp"  
FT mat\_peptide 873..1289 /\*tag- b  
FT /\*note- "contains conserved 7 cysteine skeleton"  
PN WO9304692-A.  
PD 18-MAR-1993.  
PF 28-AUG-1992; U07358.  
PR 30-AUG-1991; US-752764.  
PR 30-AUG-1991; US-752861.  
PR 30-AUG-1991; US-753059.  
PA (CREA-) CREATIVE BIOMOLECULES INC.  
PI Cohen CM, Kuberampath T, Oppermann H, Ozkaynak E;  
PI Pang RHL, Rueger DC, Smart JE;  
DR WPI: 93-100652/12.

DR N-PSDB; Q38736.  
PT Morphogen-induced modulation of inflammatory response - and  
PT resulting tissue damage, e.g. in autoimmune diseases, diabetes,  
PT asthma, ischemia reperfusion injury, etc.  
PS Claim 26: Page 126-127; 165pp; English.  
CC Osteogenic protein (OP)-2 is a preferred morphogen  
CC for use in treating tissue damage in e.g. inflammatory disease,  
CC autoimmune disease, arthritis, psoriasis, dermatitis, diabetes and  
CC emphysema. Proteins having at least 70% homology with OP-2 amino  
CC acid sequences can also be used. See R33401 for mature mop-2.  
SQ Sequence 399 AA;

Query Match 1.4%; Score 118; DB 7; Length 399;  
Best Local Similarity 38.7%; Pred. No. 4.60e+00;  
Matches 24; Conservative 12; Mismatches 22; Indels 4; Gaps 4;

Db 12 glalcalagggghppp-htc-pqrrlg-areerdmqrellavlgpgr-prpraqpaar 67  
QY 194 GPPLYQLGAATQARPPPHASGPRRRRLGGERAMNHSVREAGVPLGLPAPGARRRGSASRS 253  
Db 68 qp 69  
QY 254 LP 255

Search completed: Fri Dec 18 18:25:06 1998  
Job time : 125 secs.

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

\*\*\*\*\*  
 WISE (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch\_PP protein - protein database search, using Smith-Waterman algorithm  
 Run on: Fri Dec 18 18:20:51 1998; MasPar time 40.56 Seconds  
 971.396 Million cell updates/sec  
 Tabular output not generated.

Title: >US-08-951-733-20  
 Description: (1-1154) from US08951733.pep  
 Perfect Score: 8624  
 Sequence: 1 HASGGRCVLRTWEALAPAT.....TALEAANPALPSDFKTIID 1154

Scoring table:  
 PAM 150  
 Gap 11

Searched: 107076 seqs, 34141958 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: p1r56  
 1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 54.111; Variance 116.481; scale 0.465

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	314	3.6	884	2	SS3396	1.01e-30
2	162	1.9	660	1	QDBE3	6.63e-07
3	153	1.8	504	2	JC1306	1.12e-05
4	136	1.6	240	2	B24264	1.87e-03
5	141	1.6	276	2	B38965	4.29e-04
6	142	1.6	310	1	PIHUSD	3.19e-04
7	136	1.6	316	2	SI6881	1.87e-03
8	137	1.6	628	2	S01955	1.40e-03
9	129	1.5	264	2	D34768	1.39e-02
10	131	1.5	300	2	S19560	7.89e-03
11	126	1.5	302	2	S11790	3.23e-02
12	129	1.5	309	2	S10889	1.39e-02
13	132	1.5	317	2	A28996	5.93e-03
14	127	1.5	566	2	S22933	2.44e-02
15	132	1.5	924	2	J27923	5.93e-03
16	131	1.5	1106	2	JQ0405	7.89e-03
17	117	1.4	204	2	A39066	3.71e-01
18	125	1.4	227	2	C29149	4.26e-02
19	123	1.4	240	2	A24264	7.38e-02
20	123	1.4	256	2	A60533	7.38e-02
21	118	1.4	322	1	S00054	2.85e-01
22	119	1.4	347	2	S10571	2.18e-01
23	121	1.4	358	1	WMBE38	1.27e-01

24	125	1.4	381	2	S16506	hypothetical protein	4.26e-02
25	125	1.4	402	2	A45056	prepro osteogenic pro	7.38e-02
26	123	1.4	403	2	S2796	prip2 protein - human	4.26e-02
27	118	1.4	431	2	S09824	hypothetical protein	2.85e-01
28	123	1.4	439	2	S51939	chitinase (EC 3.2.1.1	7.38e-02
29	124	1.4	464	2	S22697	extensin - Volvox car	5.61e-02
30	119	1.4	515	2	S10572	epithelial tumor anti	2.18e-01
31	114	1.3	202	2	B36795	hypothetical protein	8.14e-01
32	115	1.3	260	2	S23373	proline-rich protein	6.28e-01
33	116	1.3	270	2	S34361	MAE protein - Salmon	4.83e-01
34	116	1.3	295	2	B48013	proline-rich proteogl	4.83e-01
35	114	1.3	301	1	E29149	proline-rich protein	8.14e-01
36	114	1.3	321	1	A28663	nodulation protein no	8.14e-01
37	115	1.3	338	2	JC5707	HVA22 protein - human	6.28e-01
38	114	1.3	373	2	S54545	hypothetical protein	8.14e-01
39	115	1.3	392	1	PIHUB6	salivary proline-rich	6.28e-01
40	114	1.3	408	2	A56186	cyclin E - African cl	8.14e-01
41	113	1.3	668	2	B35363	synapsin Ib - human	1.05e+00
42	113	1.3	704	2	A30411	synapsin Ia - rat	1.05e+00
43	114	1.3	705	2	A35363	synapsin I splice for	8.14e-01
44	113	1.3	1335	2	A35887	mucln precursor, panc	1.05e+00
45	115	1.3	3149	1	QDBE8	Bp1f1 protein - human	6.28e-01

## ALIGNMENTS

RESULT 1  
 ENTRY S53396 #type complete  
 TITLE hypothetical protein YLR318w - yeast (Saccharomyces cerevisiae)  
 ALTERNATE\_NAMES hypothetical protein L8543.12  
 ORGANISM formal name Saccharomyces cerevisiae  
 DATE 05-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 06-Feb-1998

ACCESSIONS S53396  
 REFERENCE S53390  
 #authors Du, Z.  
 #submission submitted to the EMBL Data Library, February 1995  
 #description The sequence of S. cerevisiae cosmid 8543.  
 #accession S53396

#molecule\_type DNA  
 ##residues 1-884 ##label DUZ  
 ##cross-references EMBL:U20618; NID:G2258165; PID:G662136; MIPS:YLR318w  
 ##experimental\_source strain S288C (AB972)

## GENETICS

#gene SCD:EST2  
 ##cross-references SCD:S0004310; MIPS:YLR318w  
 #map\_position 12R  
 SUMMARY #length 884 #molecular\_weight 102662 #checksum 7604

Query Match 3.6%; Score 314; DB 2; Length 884;  
 Best Local Similarity 24.3%; Pred. No. 1.01e-30;  
 Matches 118; Conservative 122; Mismatches 20; Indels 44; Gaps 37;  
 Db 271 LSHSROSPKRR-VLKFTIVYLQKLPDEMFSKKNKKIKNLILSLPLNGLYPRDS 329  
 470 LVQLRKHSSPMOYGFRAQLRLVPPGLWGSRRNRRFRANKKFLSLKHAHLSQE 529  
 Db 330 LLLKRLKDFWMLFS-DIWFTKHFNMLNOLAIC-FISWLFROLIPKTIOTFYCEIS 387  
 530 LWMKSVADCAMLRSPVGVCPAAEHRRLAKFLHMLMSYVVALMSFFVTEPT 589  
 Db 388 STVTIVF-RHDTWKKLTTPIVEYFKY-LVE-NNV-CRNHNSYTLSENHKKRIIPK 443  
 590 FOKNLFYFRKSVSKLOSIGIRHLKRVQRLSEAEVRHRE-ARPAALLTSRIRFLPK 648  
 Db 444 KSNNEFRILAIPCGADDEEFTYKENK-NA-IOPQKILEYLRNKPSTF-RTIYSPT 500  
 649 -PDGRIPIVANDY-VGKARTREKRAERLRSYKALFSLVTERARRPGIGLSVGLD 706  
 Db 501 QIADRI-KEFKORLLKKNVLPETYFKFDPVSKCYDSIPMECKR-TLKDALNENGFF 558  
 707 DI-HRAWTFVLAV-RA-QDPPPELYFKVAVTGAYDIIPQ-DRLTEVIASITIKRONTYC 762

Db	ENTRY	TITLE	ORGANISM	DATE	ACCESSIONS
Db	559	VR50Y-FENTWY-G-VLKLFVY-VNASVPPY-ELTIDNR-TVHLSNDVYV---VV-	608		
QY	763	VR-KTAVQKRAHGHVRAFKSHVSTLDDQYVNRQVYAHQETSPRLDAVIEQSSIN	821		
Db	609	E---M-EIF-K-T---ALWEDKCYIREDELFGSSLSAPIDYVDLYDDLEFYSEFKAS	658		
QY	822	EASSGLEFVFLRFMCNHAHVRIRGKSYVOCQGIPOGSIITSLTLCYCDGM-E-NKLF-AG	878		
Db	659	PSODTLTKLADDFLEISTDQOQVINTKIKLAMEGFQRYNKAN-RDKILNAVSSGDDDTV	717		
QY	879	IRRGLELRITVDDFLVTPHILTKAKTEFLRTLVKRGVPEYGCYVNRKTVNPFEDALGG	938		
Db	718	IOFCAM 723			
QY	939	TAFVQM 944			
RESULT	2				
ENTRY		QOBE3	#type complete		
TITLE		BHLF1 protein - human herpesvirus 4 (strain B95-8)			
ORGANISM		#formal name human herpesvirus 4, Epstein-Barr virus			
DATE		25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 23-Aug-1997			
ACCESSIONS		A03742			
REFERENCE		A93065			
#authors		Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.			
#journal		Mol. Biol. Med. (1983) 1:21-45			
#title		Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus.			
		#cross-references MUID:85035713			
		#accession A03742			
		#residues 1-660 #label BAN			
REFERENCE		A03794			
#authors		Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; Hatfull, G.; Hudson, G.S.; Satchwell, S.C.; Segun, C.; Tufnell, P.S.; Barrell, B.G.			
#journal		Nature (1984) 310:207-211			
#title		DNA sequence and expression of the B95-8 Epstein-Barr virus genome.			
		#cross-references MUID:84270667			
		#contents annotation; protein coding region			
COMMENT		The sequence contains four perfect repeats (residues 149-273, 274-398, 399-523, and 524-648).			
CLASSIFICATION		#superfamily human herpesvirus 4 BHLF1 protein			
SUMMARY		#length 660 #molecular_weight 66244 #checksum 8900			
Query Match		1.9%; Score 162; DB 1; Length 660;			
Best Local Similarity		30.8%; Pred. No. 6.63e-07;			
Matches		48; Conservative 34; Mismatches 60; Indels 14; Gaps 12;			
Db	251	GPPPTRGSAAG-RTHRRPCCPSARNPGCPRTWR-R-SGAKRHPPGAGQRSGP	306		
QY	194	GPPLYQCAAAQAPPPASG-PRR-R-LGGERAMNNSYRAGVPLDLPAGA-RRGGGS	249		
Db	307	TGGRPAAPGAGTGAAPGPGGGAIVSGATHPERSGPADP--AAARLPPEKQEPRLP	364		
QY	250	ASRSLPLPKRRRRGKAAPER-TRVGGGSNAHPPGRTGSPDRGCVVSPARPAEASISLE	308		
Db	365	QDLAAQRC-PAGPPPTRGSA-AAQRTNR-RPSCP	397		
QY	309	GALSGTRHSHPSVGRHAGCPPTSPRPDPDTPCP	344		
RESULT	3				
ENTRY		JC1306	#type complete		
TITLE		virion protein homolog - bovine herpesvirus 1			
ALTERNATE_NAMES		alpha T1F; BHV-1 protein homolog; ICP25; Vmw65; VP16			
ORGANISM		#formal name bovine herpesvirus 1			
DATE		05-Mar-1993 #sequence_revision 05-Mar-1993 #text_change 09-Sep-1997			
ACCESSIONS		JC1306; S24229			

REFERENCE	JC1306
#author	Carpenter, D.E.; Mistr, V.
#journal	Gene (1992) 119:259-263
#title	Sequences of the bovine herpesvirus 1 homologue of herpes simplex virus type-1 alpha-trans-inducing factor (UL48).
#accession	JC1306
##molecule_type	DNA
##residues	1-504 ##label CAR
##cross-references	EMBL:J11610; NID:g1065725; PID:e64419; PID:g1065726
COMMENT	This protein interacts with cellular transcription factors to transactivate immediate early viral genes.
GENETICS	
KEYWORDS	map_position 0.07-0.086
SUMMARY	DNA binding; transcription regulation
Query Match	1.8%; Score 153; DB 2; Length 504;
Best Local Similarity	31.3%; Pred. No.1,12e-05;
Matches	46; Conservative 33; Mismatches 56; Indels 12; Gaps 11;
Db	351 APABGGGWRSSGSTRGRGAANSTIGRLQRCGCCRRRRAC-CHAF-RORLR--ARGE 406
Qy	208 PPPHASPRRRLCCERAMNHSVREACVPGLDPAPARRGGSASASLPLPRPRGAPE 267
Db	407 PRRHS-GSGAFSSQ-GRPPGVLCRLGACAKARSGPAGGSPVPSGLGSR-ARGSPGP 463
Qy	268 PERFPVQGSWAHGRTRGSPDR-GR-CVY-S-PARPAEATSLGALSGTRHSHPSYGR 323
Db	464 GPACGSPSRARGGRRRASPANP-FGGT 489
Qy	324 QHHAGPPSTSRPPRPMDTPCPYAT 350
RESULT	4
ENTRY	B24264
TITLE	proline-rich protein MP3 - mouse (fragment)
ORGANISM	#formal name Mus musculus #common name house mouse
DATE	09-Sep-1987 #sequence_revision 09-Sep-1987 #text_change 03-May-1996
ACCESSIONS	B24264
REFERENCE	A92508
#authors	Ann, D.K.; Carlson, D.M.
#journal	J. Biol. Chem. (1985) 260:15863-15872
#title	The structure and organization of a proline-rich protein gene of a mouse multi-gene family.
#cross-references	MJID:86059475
#accession	B24264
##molecule_type	DNA
##residues	1-240 ##label ANN
CLASSIFICATION	#superfamily proline-rich protein
SUMMARY	#length 240 #checksum 5152
Query Match	1.6%; Score 136; DB 2; Length 240;
Best Local Similarity	27.2%; Pred. No.1,87e-03;
Matches	41; Conservative 38; Mismatches 65; Indels 7; Gaps 6;
Db	24 VNGSQGPPPPGGGPPRRPPGCGPPGPPRRPPGPPGPPGPPGPPGPPGPPR 83
Qy	200 LGATQARPPHASPGRRLGGERAMNHSVREAGVPLGLPAPGARRRGSGASRSLPLPKR 259
Db	84 PQGPPPPGGGQQRP-PQGPPPPGGGPPRRPPGCGPPPPAGGPPPPGPPGPHLRPQG 142
Qy	260 PRRGAPE-PEFRPVQGSWAHGR-RTGRPSDRGFCVVSPPARPAEATSLGCA-LSGTRH 316
Db	143 PPPGSPQRPQGGPPPTGPPPTGPP-PP 172
Qy	317 SHPSVGKQHHN--GPPSTSRPPRPMDTPCP 345
RESULT	5
ENTRY	B38965
TITLE	hypothetical protein B (insertion sequence IS1222) - Enterobacter agglomerans

```

REFERENCE      A38355
#authors       Kauffman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.
#journal       Biochemistry (1991) 30:3351-3356
#title         Basic proline-rich proteins from human parotid saliva:
               relationships of the covalent structures of ten proteins
               from a single individual.
#cross-references MIMD:91190864
#accession     E38355
#molecule-type protein
##residues    241-254,'KN',257-310 ##label KAU
REFERENCE      A03295
#authors       Satoh, E.; Isemura, S.; Sanada, K.
#journal       J. Biochem. (1983) 93:495-502
#title         Complete amino acid sequence of a basic proline-rich peptide
               P-D, from human parotid saliva.
#cross-references MIMD:83186122
#accession     A03295
#molecule-type protein
##residues    241-310 ##label SAI
REFERENCE      A61294
#authors       Shimomura, H.; Kanai, Y.; Sanada, K.
#journal       J. Biochem. (1983) 93:857-863
#title         Amino acid sequences of glycopeptides obtained from basic
               proline-rich glycoprotein of human parotid saliva.
#accession     A61294
#molecule-type protein
##residues    54-57,'E','59-73','R':82-101 ##label SHI
REFERENCE      C62891
#authors       Charlton, A.J.; Baxter, N.J.; Lilley, T.H.; Haslam, E.;
               McDonald, C.J.; Williamson, M.P.
#journal       FEBS Lett. (1996) 382:289-292
#title         Tannin interactions with a full-length human salivary
               proline-rich protein display a stronger affinity than with
               single proline-rich repeats.
#accession     S62891
#molecule-type protein
##residues    241-252 ##label CHA
               amino end of peptide designated basic proline-rich
               protein IB-5
#note          It is unclear from the peptide sequence whether this is
               a product of the PRB2 (P.R:P.HUPEF) or PRB4 (this
               entry) gene
GENETICS
#gene          GDB:A38355
               #cross-references GDB:119514; OMIM:180990
               #map-position 12p13.2-12p13.2
               #introns   22/1: 34/1
               #note          The list of introns may be incomplete
               CLASSIFICATION #superfamily proline-rich protein
               KEYWORDS    glycoprotein; saliva; tandem repeat
FEATURE
1-16           #domain signal sequence #status predicted #label SIG\
241-310        #product proline-rich peptide P-D #status experimental
               #binding-site carbohydrate (Asn) (covalent) #status
               experimental\
66,87,171      #binding-site carbohydrate (Asn) (covalent) #status
               predicted
108,150,192,213,
234            #binding-site carbohydrate (Asn) (covalent) #status
               predicted
SUMMARY        #length 310 #molecular-weight 31351 #checksum 3960
Query Match   1.6%; Score 142; DB 1; Length 310;
Best Local Similarity 28.2%; Pred. No. 3.19e-04;
Matches       40; Conservative 41; Mismatches 52; Indels 9; Gaps 9;
Db           168 QGQNSOG-PPHNGKRRPP-QGG-NQSHRRPPP-GPER-PPPGGNOSOG-P-PP 220
              | :|: ||||: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: |
Oy           199 QLGATARPAPHASGRRRLGGERANHSVRAGVPLGLPADGARRRGSASRSJLPLP 258
              | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: |
Yy           259 RPRGAAPRPRTIVGGGSAWHNRKTRGPDR-GFCVSWSPARPAE-EATSLDGALSGTKH 316

```

D6	281	APAGKPGGPPRPGGRPRP	302
OY	317	SHPVSGROHAGPPTSRPRP	338
RESULT	7		
ENTRY		S16681	#type complete
TITLE		homeotic protein - human	
ORGANISM		21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 17-Oct-1997	
DATE			
ACCESSIONS		S16681	
REFERENCE		S16681	
#authors		Deaguchi, Y.; Kehrl, J.H.	
#journal		Nucleic Acids Res. (1991) 19:3742	
#title		Nucleotide sequence of a novel diverged human homeobox gene encodes a DNA binding protein.	
#cross-references MUID:		91305125	
#accession		S16681	
#status		preliminary; nucleic acid sequence not shown;	
		translation not shown	
##molecule_type		mRNA	
##residues		1-316 ##label DEG	
##cross-references EMBL:		X56537	
#note		the nucleotide sequence was submitted to the EMBL Data Library, January 1991	
CLASSIFICATION		#superfamily unassigned homeobox proteins; homeobox homology DNA binding; homeobox; nucleus; transcription regulation	
KEYWORDS			
FEATURE			
73-127			
SUMMARY		#domain homeobox homology #label HOX	
		#length 316 #molecular_weight 34713 #checksum 3095	
Query Match		1.6%; Score 136; DB 2; Length 316;	
Best Local Similarity		30.9%; Pred. No. 1.87e-03;	
Matches		21; Conservative 20; Mismatches 24; Indels 3; Gaps 3;	
D6	202	RPMSTASDCVGATGTGTPRGPAI-SRPSPSSPAAQDRPARAKAPGPASPGLAWTH	260
OY	223	RAWHNSVEAEV-PLGIPIAPARRRRGSASSRLPLPRPRRG-AAPEPERTPVGGGSMW	280
D6	261	PARPDEA	268
OY	281	PGRTGPS	288
RESULT	8		
ENTRY		S01955	#type complete
TITLE		hypothetical protein, 69k - turnip yellow mosaic virus	
ORGANISM		#formal_name turnip yellow mosaic virus, TYMV	
DATE		21-Nov-1993 #sequence_revision 26-May-1995 #text_change 26-May-1995	
ACCESSIONS		S01955	
REFERENCE		S01955	
#authors		Morch, M.D.; Boyer, J.C.; Haenni, A.L.	
#journal		Nucleic Acids Res. (1988) 16:6157-6173	
#title		Overlapping open reading frames revealed by complete nucleotide sequencing of turnip yellow mosaic virus genomic RNA.	
#cross-references MUID:		88289359	
#accession		S01955	
#status		preliminary	
##molecule_type		genomic RNA	
##residues		1-628 ##label MOR	
##cross-references EMBL:		X07441	
#note		the authors translated the codon ACG for residue 459 as U	
SUMMARY		#length 628 #molecular_weight 69194 #checksum 7569	
Query Match		1.6%; Score 137; DB 2; Length 628;	
Best Local Similarity		25.0%; Pred. No. 1.40e-03;	
Matches		73; Conservative 64; Mismatches 134; Indels 21; Gaps 20;	
D6	157	GPIVLTKPKRTSVQBPBSATRGPSFRILLPKV-VH-VDDPHSSLRPRSRSRLQDPT	214

Qy	194	GPPLYLGATQAQAPRPNHAS - GPR - KRLGCEBRAMNHSVRRAGVPLGIPAGCAR - RR - GGS 249
Db	215	VRRPLAQNPHQSHSRQDPPPLSDPDGILGPPPLAHPHSTRDDPPRI - TPQPSN - THDLRPL 272
Qy	250	ASRSLPLPKRRRRCGAADPEPTPVG - QGSNA - HPGRTGRPSDQFCVSPAPRAEATSL 307
Db	273	SVLPRTSRRKLLNPNRRHRTSTGHIPPTTTSRPTGSPSKLQRPVH - LYOSSPHTPFRP 331
Qy	308	EG - ALSGRHSHPVGVGROHNAAGPSTSRPPRPWPTDPPPYAETKRLV - SSQDKRQLAP 365
Db	332	SSIRKDALQCPGLGHLERLGCQANLRTSERSRPTKRLPRSESEPRLKPLPEATLAP 391
Qy	366	SFLI - SSI - R - PELTGARKLVEITIFGSRPMWGTFRRLPRLPQRWKQRPFLLELG - 420
Qy	421	NRHQC - PYGVLLKTHCPRLRA - AVTPAAGVOCAREKPDQGSVAAPEDDTPDRL 470
RESULT	9	
ENTRY	D34768	#type complete
TITLE	ORF4 protein - Orf virus (strain Nz2)	
ORGANISM	#formal_name Orf virus	
DATE	23-Aug-1991	#sequence_revision 23-Aug-1991
	31-Oct-1997	#text_change
ACCESSIONS	D34768	
REFERENCE	A34768	
#authors	Fraser, K.M.; Hall, D.F.; Mercer, A.A.; Robinson, A.J.	
#journal	Virology (1990) 176:370-389	
#title	Sequence analysis of the inverted terminal repetition in the genome of the parapoxvirus, orf virus.	
#cross-references	MUID:90266454	
#accession	D34768	
##status	Preliminary	
##molecule_type	DNA	
##residues	1-264	##label FRA
##cross-references	GB:M30023; EMBL:M37623; NID:g9332561; PID:g9332566	
SUMMARY	#length 264	#molecular_weight 25613
		#checksum 1957
Query Match	1.5%	Score 129; DB 2; Length 264;
Best Local Similarity	28.2%	Pred. No.1,39e-02;
Matches	33; Conservative	40; Mismatches 37; Indels 7; Gaps 5;
Db	150	RAPGRAGRYPARAPAAASAAASRRHRGPAARSSPFAERPPAPPARAEAPARRASS - GS 208
Qy	223	RAMNHSVR - EAGVPLGLDIPAPGARRRGGSASRSLPKRP - - - R - RGAADPEPTPVQGS 277
Db	209	RASSGPPRRSAARS - SAASGSRPAASGPAPAPAPAPASSARTSAGEGAARGCAPACW 264
Qy	278	WAHPGRTGRGPDRCFCVYSPAPRAEEATSLGALSGRHSHPVGRQHNAGPPSTSR 334
RESULT	10	
ENTRY	S19560	#type complete
TITLE	proline-rich protein MP4 - mouse	
ORGANISM	#formal_name Mus musculus	#common_name house mouse
DATE	22-Nov-1993	#sequence_revision 10-Nov-1995
	20-Mar-1998	#text_change
ACCESSIONS	S19560; S22570	
REFERENCE	S19560	
#authors	Roberts, S.G.E.; Layfield, R.; Bannister, A.J.; McDonald, C.J.	
#journal	Eur. J. Biochem. (1991) 202:969-974	
#title	Gene sequence of mouse B-type proline-rich protein MP4. Transcriptional start point and an upstream phylogenetic footprint with ets-like and rel/NFkB-like elements.	
#cross-references	MUID:92111548	
#accession	S19560	
##status	Preliminary	
##molecule_type	DNA	
##residues	1-300	##label ROB
##cross-references	GB:X58438; NID:g953181; PID:g953182	



Search completed: Fri Dec 18 18:22:43 1998  
Job time : 112 secs.

\*\*\*\*\*  
 WISE (TM)  
 \*\*\*\*\*

Release 3.1a John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Fri Dec 18 18:25:25 1998; MasPar time 30.89 Seconds  
 Tabular output not generated. 1002.720 Million cell updates/sec

Title: >US-08-951-733-20  
 Description: (1-1154) from US08951733.pep  
 Perfect Score: 8624  
 Sequence: 1 HNSGQRCVLRTWEALAPAT.....TALEAANPALPSDFITLD 1154

Scoring table: PAM 150  
 Gap 11

Searched: 74019 segs, 26840295 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: swiss-prot35  
 1:swissprot

Statistics: Mean 55.898; Variance 96.260; scale 0.581

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	162	1.9	560	1	YH1_EBV	4.64e-09
2	153	1.8	504	1	ATIN_HSVBP	1.47e-07
3	150	1.7	234	1	PRPM_HUMAN	4.56e-07
4	142	1.6	276	1	PRPL_HUMAN	8.72e-06
5	137	1.6	628	1	V70K_TYMV	5.26e-05
6	133	1.5	247	1	PRP4_HUMAN	2.15e-04
7	131	1.5	296	1	PRP3_MOUSE	4.31e-04
8	127	1.5	302	1	NOB3_RHILP	2.37e-03
9	126	1.5	566	1	TS13_MOUSE	1.69e-03
10	125	1.4	261	1	PRP2_MOUSE	3.32e-03
11	121	1.4	322	1	NOB1_RHIN	1.25e-02
12	118	1.4	399	1	BMB4_MOUSE	3.30e-02
13	115	1.4	402	1	BMP8_HUMAN	3.32e-03
14	118	1.4	431	1	UL61_HCMVA	3.30e-02
15	125	1.4	501	1	MEF1_HUMAN	3.32e-03
16	118	1.4	514	1	VE2_HPV05	3.30e-02
17	114	1.3	202	1	VG01_HSVB	1.17e-01
18	116	1.3	270	1	MIAE_SALTY	6.22e-02
19	115	1.3	279	1	Y091_NPVOP	8.54e-02
20	109	1.3	301	1	ND03_RHIL	5.40e-01
21	114	1.3	321	1	NOB1_RHIL	1.17e-01
22	109	1.3	324	1	MATD_NEUCR	5.40e-01
23	115	1.3	331	1	PRP1_HUMAN	8.54e-02

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
24	114	1.3	408	1	CGE1_XENLA	1.17e-01
25	114	1.3	408	1	CGE2_XENLA	1.17e-01
26	110	1.3	408	1	G1/S-SPECIFIC CYCLIN E	3.99e-01
27	111	1.3	498	1	VE2_XENLA	2.95e-01
28	111	1.3	509	1	VE2_HPV08	2.95e-01
29	113	1.3	514	1	VE2_HPV36	1.59e-01
30	110	1.3	628	1	VE2_HPV5B	3.99e-01
31	112	1.3	633	1	VE2_HPV5B	2.17e-01
32	113	1.3	704	1	VE2_HPV5B	1.59e-01
33	114	1.3	705	1	VE2_HPV5B	1.17e-01
34	108	1.3	706	1	VE2_HPV5B	7.28e-01
35	113	1.3	1255	1	VE2_HPV5B	1.59e-01
36	115	1.3	1349	1	VE2_HPV5B	8.54e-02
37	107	1.2	174	1	VE2_HPV5B	9.79e-01
38	106	1.2	233	1	VE2_HPV5B	1.31e+00
39	106	1.2	285	1	VE2_HPV5B	1.31e+00
40	107	1.2	318	1	VE2_HPV5B	9.79e-01
41	107	1.2	494	1	VE2_HPV5B	9.79e-01
42	107	1.2	529	1	VE2_HPV5B	9.79e-01
43	105	1.2	863	1	VE2_HPV5B	1.76e+00
44	105	1.2	890	1	VE2_HPV5B	1.76e+00
45	105	1.2	2774	1	VE2_HPV5B	1.76e+00

## ALIGNMENTS

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	162	1.9	560	1	YH1_EBV	4.64e-09
2	153	1.8	504	1	ATIN_HSVBP	1.47e-07
3	150	1.7	234	1	PRPM_HUMAN	4.56e-07
4	142	1.6	276	1	PRPL_HUMAN	8.72e-06
5	137	1.6	628	1	V70K_TYMV	5.26e-05
6	133	1.5	247	1	PRP4_HUMAN	2.15e-04
7	131	1.5	296	1	PRP3_MOUSE	4.31e-04
8	127	1.5	302	1	NOB3_RHILP	2.37e-03
9	126	1.5	566	1	TS13_MOUSE	1.69e-03
10	125	1.4	261	1	PRP2_MOUSE	3.32e-03
11	121	1.4	322	1	NOB1_RHIN	1.25e-02
12	118	1.4	399	1	BMB4_MOUSE	3.30e-02
13	115	1.4	402	1	BMP8_HUMAN	3.32e-03
14	118	1.4	431	1	UL61_HCMVA	3.30e-02
15	125	1.4	501	1	MEF1_HUMAN	3.32e-03
16	118	1.4	514	1	VE2_HPV05	3.30e-02
17	114	1.3	202	1	VG01_HSVB	1.17e-01
18	116	1.3	270	1	MIAE_SALTY	6.22e-02
19	115	1.3	279	1	Y091_NPVOP	8.54e-02
20	109	1.3	301	1	ND03_RHIL	5.40e-01
21	114	1.3	321	1	NOB1_RHIL	1.17e-01
22	109	1.3	324	1	MATD_NEUCR	5.40e-01
23	115	1.3	331	1	PRP1_HUMAN	8.54e-02

01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)  
01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)  
ALPHA TRANS-INDUCING PROTEIN (ALPHA-TIF).  
BOVINE HERPESVIRUS TYPE 1 (STRAIN PB-2).  
VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; ALPHAHERPESVIRINAE.  
[1]  
SEQUENCE FROM N.A.  
MEDLINE: 93012995.  
RA CARPENTER D.E., MISRA V.;  
RL GENE 119:259-263(1992).  
CC - FUNCTION: RESPONSIBLE FOR TRANSCRIPTIONAL ACTIVATION OF IMMEDIATE  
EARLY PROMOTERS (ALPHA GENES) (BY SIMILARITY).  
CC - SIMILARITY: TO OTHER HERPESVIRUSES ALPHA TRANS-INDUCING PROTEIN.  
DR EMBL: Z11610; E264419; -.  
DR PIR: S24229; S24229.  
DR PIR: JC1306; JC1306.  
KW TRANSCRIPTION REGULATION; TRANS-ACTING FACTOR; DNA-BINDING.  
SQ SEQUENCE 504 AA; 54028 MW; 79F42020 CRC32;  
Query Match 1.8%; Score 153; DB 1; Length 504;  
Best Local Similarity 31.3%; Pred. No. 1.47e-07;  
Matches 46; Conservative 33; Mismatches 56; Indels 12; Gaps 11;  
Db 351 APEAGGGRGSTRTRGRARSTGRLOPCGPRRRAC-CRATP-RORLR--ARGE 406  
Qy 208 PPPASGRRRLGRCERAMNHSYREAGVPLGLPAPARRGGSSASLSPLPKRRGAPE 267  
Db 407 PRHTS-GSGAESQ-GRPGRCRLGMACARSGPARGGSPVNSGLSLF-ARGSPG 463  
Qy 268 PERTVGGGSMHAPGRTGSPDR-GF-CVV-S-PAPAEATSLGALSGTRHSPVGR 323  
Db 464 GPACGPRRARGRRRASPANP-FGGT 489  
Qy 324 OHNAGPSTSRPPRWDPPCPVYAE 350  
RESULT 3  
ID PRPM\_HUMAN STANDARD; PRT; 234 AA.  
AC P10161; P02813;  
DT 01-MAR-1989 (REL. 10, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE SALIVARY PROLINE-RICH PROTEIN PO (ALLELE M) (CONTAINS: PEPTIDE P-D)  
DE (FRAGMENT).  
GN PRB4.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 89121439.  
RA LYONS K.M., STEIN J.H., SMITHIES O.;  
RL GENETICS 120:255-265(1988).  
RN [2]  
RP SEQUENCE OF 165-234.  
RX MEDLINE: 83186122.  
RA SALTON E., ISEMURA S., SANADA K.;  
RL J. BIOCHEM. 93:495-502(1983).  
DR EMBL: X07704; E265547; -.  
DR PIR: A03295; PIHUSD.  
DR PIR: S03175; S03175.  
DR HSSP: P19999; ICLG.  
DR MIM: 168730; -.  
DR MIM: 180990; -.  
KW REPEAT; PAROTID GLAND; SALIVA; MULTIGENE FAMILY.  
FT NON\_TER 1  
FT CHAIN 165 234 PEPTIDE P-D.  
FT SEQUENCE 234 AA; 23676 MW; EDD24ADC CRC32;  
Query Match 1.7%; Score 150; DB 1; Length 234;  
Best Local Similarity 29.3%; Pred. No. 4.56e-07;  
Matches 43; Conservative 41; Mismatches 53; Indels 10; Gaps 10;

88 GPP-OGGNSOG-PPHPGKRRPP-OGG-NOSHRPPPP-GKPER-PPPOGNSOG 141  
Qy 194 GPPYQGAQAQAPPPASPRRLCCERAMNHSYREAGVPLGLPAPARRGGSSASLS 253  
Db 142 -P-PPHPGKRRPP-OGGNSOG-SRARSPPGKPOGPOGNGKPPGPPGKPOGPPPGN 199  
Qy 254 LPLPKRRPGAAPEPERTVGGGSMHAPGRTGSPDR-GFCVSPAPAE-EATSLGAL 311  
Db 200 PQOPAPPPAGKPPGPPPPGKRRPP 226  
Qy 312 SGTNRHSPSVGRQHHAGPSTSRPPR 338  
RESULT 4  
ID PRPL\_HUMAN STANDARD; PRT; 276 AA.  
AC P10162; P02813;  
DT 01-MAR-1989 (REL. 10, CREATED)  
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)  
DE SALIVARY PROLINE-RICH PROTEIN PO (ALLELE K) (CONTAINS: PEPTIDE P-D)  
DE (FRAGMENT).  
GN PRB4.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 89121439.  
RA LYONS K.M., STEIN J.H., SMITHIES O.;  
RL GENETICS 120:255-265(1988).  
RN [2]  
RP SEQUENCE OF 207-276.  
RX MEDLINE: 83186122.  
RA SALTON E., ISEMURA S., SANADA K.;  
RL J. BIOCHEM. 93:495-502(1983).  
DR EMBL: X07715; E4806; ALT\_SEQ.  
DR PIR: A03295; PIHUSD.  
DR PIR: S03176; S03176.  
DR HSSP: P19999; ICLG.  
DR MIM: 168730; -.  
DR MIM: 180990; -.  
KW REPEAT; PAROTID GLAND; SALIVA; MULTIGENE FAMILY.  
FT NON\_TER 1  
FT CHAIN 207 276 PEPTIDE P-D.  
FT SEQUENCE 276 AA; 27816 MW; 4838945A CRC32;  
Query Match 1.6%; Score 142; DB 1; Length 276;  
Best Local Similarity 28.2%; Pred. No. 8.72e-06;  
Matches 40; Conservative 41; Mismatches 52; Indels 9; Gaps 9;  
Db 134 QGGNSOG-PPHPGKRRPP-OGG-NOSHRPPPP-GKPER-PPPOGNSOG-P-PP 186  
Qy 199 QLGATQARPPHSGPRRLGRCERAMNHSYREAGVPLGLPAPARRGGSSASLSPLK 258  
Db 187 HPKREPPGPPGKRRSARSPPGKPOGPOGNGKPPGPPGKPPGPPGPPGPPGPP 246  
Qy 259 PPRGAPEPERTVGGGSMHAPGRTGSPDR-GFCVSPAPAE-EATSLGALSGTRH 316  
Db 247 APPAGKPPGPPPPGKRRPP 268  
Qy 317 SHPSVGRQHHAGPSTSRPPR 338  
RESULT 5  
ID V7OK\_TYMV STANDARD; PRT; 628 AA.  
AC P10357;  
DT 01-MAR-1989 (REL. 10, CREATED)  
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)  
DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)  
DE 69 KD PROTEIN.  
OS TURNIP YELLOW MOSAIC VIRUS.  
OC VIRIDAE; SS-RNA NONENVELOPED VIRUSES; TYMOVIRIDAE.  
RN [1]



CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL  
 CC REGULATORS.  
 DR EMBL: X54215; G46237; -.  
 DR PIR: S11790; S11790.  
 DR PROSITE: PS00044; HTH\_LYSR\_FAMILY; 1.  
 KW MODULATION; TRANSCRIPTION REGULATION; DNA-BINDING; ACTIVATOR;  
 KW REPRESSOR; PLASMID.  
 FT DNA\_BIND 23  
 SQ SEQUENCE 302 AA; 34274 MW; ED62D2FA CRC32;  
 Query Match 1.5%; Score 126; DB 1; Length 302;  
 Best Local Similarity 31.4%; Pred. No. 2.37e-03;  
 Matches 38; Conservative 28; Mismatches 46; Indels 9; Gaps 9;  
 Db 17 LMIEENLTAAARSTLSPAMSAVRL-RSTFDEL-FTMRGEFTVP-R-AEDLAPA 72  
 Oy 9 LLRTWEALAPATP-AMPAPRCRAVSLRSHYREVLPALFVRRLGQGWLVORGPDA 67  
 Db 73 IREALQIRLNIIPMDKFTPDSDRHR-VS-ICDEFVTVLEFKILERLAREAPGISFDL 130  
 Oy 68 AFRALVACLVCPWDA-RPPAPASFQVQSCLELVARVL-QRLCEGAKNVLAFFGRL 125  
 Db 131 L 131  
 Oy 126 L 126  
 RESULT 9  
 ID TS13\_MOUSE STANDARD; PRT; 566 AA.  
 AC Q01755;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE TESTIS-SPECIFIC PROTEIN PBS13.  
 GN TCP11.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUPAROTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN EUHERIA; RODENTIA.  
 RP [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CBA/CA; TISSUE-TESTIS;  
 RX MEDLINE; 91372153.  
 RA MACARAKIS N.D., NELKI D., LYON M.F., EVANS E.P., RUDDY S.,  
 RA FREEMONT P., DUDLEY K.;  
 RL DEVELOPMENT 111:561-571(1991).  
 CC -1- FUNCTION: POSSIBLY PLAYS AN IMPORTANT ROLE IN SPERM DEVELOPMENT  
 CC AND FUNCTION.  
 CC -1- TISSUE SPECIFICITY: TESTIS.  
 CC -1- DEVELOPMENTAL STAGE: FIRSTLY EXPRESSED IN THE PACHYTENE  
 CC SPERMATOCYTE STAGE.  
 CC -1- SIMILARITY: SOME TO YEAST SOK1.  
 DR EMBL: X52128; G54853; -.  
 DR PIR: S22933; S22933.  
 DR MGI: S22933; S22933.  
 DR MGD: MGI:98544; TCP11.  
 KW TESTIS; SPERMATOGENESIS; REPEAT; COILED COIL.  
 FT DOMAIN 171 206  
 FT DOMAIN 256 318  
 FT DOMAIN 419 432  
 FT DOMAIN 440 509  
 FT DOMAIN 298 318  
 FT REPEAT 298 304  
 FT REPEAT 305 311  
 FT REPEAT 312 318  
 FT REPEAT 318 318  
 SQ SEQUENCE 566 AA; 61970 MW; 7A421F03 CRC32;  
 Query Match 1.5%; Score 127; DB 1; Length 566;  
 Best Local Similarity 35.7%; Pred. No. 1.69e-03;  
 Matches 30; Conservative 19; Mismatches 30; Indels 5; Gaps 5;  
 Db 34 SARGIDRVSTVARAPSPGPRGAVKTAPRGVGHGLTGTSTRCPOPASARAKLPS 93  
 Oy 239 PAPGARRRGSGASRLPL-PRKPRGALPEPERTVGGGS-WAHP-GTRRGSDRG-FCV 294

Db 94 VTRGAPLPSPG-KGHLGTPSSH 116  
 Oy 295 VSPAPAEATSLGALSGTRHSH 318  
 RESULT 10  
 ID PRP2\_MOUSE STANDARD; PRT; 261 AA.  
 AC P05142;  
 DT 13-AUG-1987 (REL. 05, CREATED)  
 DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)  
 DE PROLINE-RICH PROTEIN MP-2 PRECURSOR.  
 GN PRP.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUPAROTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN EUHERIA; RODENTIA.  
 RP [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 86059475.  
 RA ANN D.K., CARLSON D.M.;  
 RL J. BIOL. CHEM. 260:15863-15872(1985).  
 DR EMBL: M12099; G200547; -.  
 DR HSSP; P19999; ICLG.  
 KW REPEAT; SALIVA; SIGNAL.  
 FT SIGNAL 1 15  
 FT CHAIN 16 261  
 SQ SEQUENCE 261 AA; 26034 MW; 9D830DAF CRC32;  
 Query Match 1.4%; Score 125; DB 1; Length 261;  
 Best Local Similarity 27.0%; Pred. No. 3.32e-03;  
 Matches 40; Conservative 33; Mismatches 69; Indels 6; Gaps 5;  
 Db 76 PRPGPPPGGQPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPG 135  
 Oy 203 ATOARPPPHASGPRRRRCERAMNHSVREAGVPLGLPAPGARRRGSGASRLPLKPRR 262  
 Db 136 PRPGGPPOLRRPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPP 194  
 Oy 263 GAPE-PE-RTPVGGSGMAHGRTRGSPSDRGFCVSPARAEATSLGALSGTRHSHPS 320  
 Db 195 GGPQQPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPG 222  
 Oy 321 VGRQHHA--GPPSTSRP-PRWDPICPP 345  
 RESULT 11  
 ID NOD1\_RHISN STANDARD; PRT; 322 AA.  
 AC P55359;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE MODULATION PROTEIN D I.  
 GN NODDI OR Y4AL.  
 OS RHIZOBIDUM SP. (STRAIN NGR234).  
 OG PLASMID SYM PNGR234A.  
 CC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;  
 CC RHIZOBIACEAE.  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 RP MEDLINE; 97305956.  
 RX FREIBERG C.A., FELLAY R., BAIRROCH A., BROUGHTON W.J., ROSENTHAL A.,  
 RA PERRET X.;  
 RL NATURE 387:394-401(1997).  
 CC -1- FUNCTION: NODD REGULATES THE EXPRESSION OF THE NOD ABCFE GENES  
 CC WHICH ENCODE OTHER MODULATION PROTEINS. NODD IS ALSO A NEGATIVE  
 CC REGULATOR OF ITS OWN EXPRESSION. BINDS FLAVONOIDS AS INDUCERS.  
 CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL  
 CC REGULATORS.  
 DR EMBL: A600065; G2182301; -.  
 DR PROSITE: PS00044; HTH\_LYSR\_FAMILY; 1.  
 KW MODULATION; TRANSCRIPTION REGULATION; DNA-BINDING; ACTIVATOR;  
 KW REPRESSOR; MULTIGENE FAMILY; PLASMID.  
 FT DNA\_BIND 23  
 SQ SEQUENCE 322 AA; 34274 MW; ED62D2FA CRC32;  
 Query Match 1.5%; Score 127; DB 1; Length 322;  
 Best Local Similarity 35.7%; Pred. No. 1.69e-03;  
 Matches 30; Conservative 19; Mismatches 30; Indels 5; Gaps 5;  
 Db 34 SARGIDRVSTVARAPSPGPRGAVKTAPRGVGHGLTGTSTRCPOPASARAKLPS 93  
 Oy 239 PAPGARRRGSGASRLPL-PRKPRGALPEPERTVGGGS-WAHP-GTRRGSDRG-FCV 294



SEQ SEQUENCE 431 AA; 44309 MM; 232AB9D7 CRC32;

Query Match 1.4%; Score 118; DB 1; Length 431;

Best Local Similarity 32.1%; Pred. No. 3.30e-02;

Matches 27; Conservative 20; Mismatches 31; Indels 6; Gaps 6;

Db 268 RGGGKPPGLSPRA-TDNRDPC-A-GVAPRRRRGSSGGGTPGGRPERAAGAP 324

QY 211 HASGPRRLGCCRANMNSVREAGVPLGLPAPARRRGSSAS-RSLPLPKRRRGAAPPE 269

Db 325 TAPDAPGRWDGPDADGAPAGLG 348

QY 270 RTPVG-OGSWAPGRTRGPPS-DRG 291

RESULT 15

ID MEH1\_HUMAN STANDARD; PRT; 501 AA.

AC Q99958;

DT 15-JUL-1998 (REL. 36, CREATED)

DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)

DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE MESENCHYME FORK HEAD PROTEIN 1 (MEH-1 PROTEIN) (TRANSCRIPTION FACTOR

FKH-14).

GN FKH14 OR MEH1.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUHERIA; PRIMATES.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 97312712.

RA MIURA N., IIDA K., KAKINUMA H., YANG X.-L., SUGIYAMA T.;

RL GENOMICS 41:489-492(1997).

CC -1- FUNCTION: MIGHT BE INVOLVED IN THE FORMATION OF SPECIAL

MESENCHYMAL TISSUES (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).

CC EMBL; Y08223; E303016; -.

DR MIM; 602402; -.

DR PROSITE; PS00657; FORK\_HEAD\_1; 1.

DR PROSITE; PS00658; FORK\_HEAD\_2; 1.

DR PROSITE; PS50039; FORK\_HEAD\_3; 1.

KW TRANSCRIPTION REGULATION; DNA-BINDING; NUCLEAR PROTEIN.

FT DNA\_BIND 71 162 FORK-HEAD.

FT DOMAIN 163 167 POLY-ARG.

FT DOMAIN 387 396 HIS-RICH.

FT DOMAIN 397 421 ALA/PRO-RICH.

FT DOMAIN 400 408 POLY-PRO.

FT DOMAIN 416 422 POLY-ALA.

SEQ SEQUENCE 501 AA; 53719 MM; 99D32EFA CRC32;

Query Match 1.4%; Score 125; DB 1; Length 501;

Best Local Similarity 28.4%; Pred. No. 3.32e-03;

Matches 27; Conservative 28; Mismatches 38; Indels 2; Gaps 2;

Db 367 SPLSLNTLAOCGALATGHHHHPQAPPPAPQPPQPGCAAAQAASWYL 426

QY 296 SPAREAEATSTEGALSGTRHSHPSVGRQH-HAGPPSTSRPPRPMDTQCPVYATETKFL 354

Db 427 NMSGDLNHL-PGHTFAAOQGTFFPNVREMFSRLG 460

QY 355 YSSGDKQLRPSFLSLRPSLTGARLVETITLG 389

Search completed: Fri Dec 18 18:26:40 1998  
Job time : 75 secs.



```

OY 263 GAAPERPTVPGGSMWHPGRTGSPDGRGFCVVSAPARPAEATSELEGALSGTRSHSPSVG 322
Db 301 RQNHAGPSTSPRPMPGTPCPRPVYAEKHFILYSSGDKQQLRPSFLSSLSRSLTGARL 360
OY 323 RQNHAGPSTSPRPMPGTPCPRPVYAEKHFILYSSGDKQQLRPSFLSSLSRSLTGARL 382
Db 361 VETIFLGRPMWPGTPRRLPRLPQRYWOMRPLFELLGNHACCPGYLKTCPRLAAVT 420
OY 383 VETIFLGRPMWPGTPRRLPRLPQRYWOMRPLFELLGNHACCPGYLKTCPRLAAVT 442
Db 421 PAAGCAREKPGQSVAAABEEDTDPRRLVQLLRQSSPMQYGVYFACLRRLVPPGLWS 480
OY 443 PAAGCAREKPGQSVAAABEEDTDPRRLVQLLRQSSPMQYGVYFACLRRLVPPGLWS 502
Db 461 RHNERFLRNTKFTISLQKHAHLSLQELTKMSYVDCAMLRSPGVGCYPAAEHLRBEI 540
OY 503 RHNERFLRNTKFTISLQKHAHLSLQELTKMSYVDCAMLRSPGVGCYPAAEHLRBEI 562
Db 541 LAKFLHMLMSYVVELLSFEYVETTFQKNRLEFYRKSVMKLSIGIRQHLKRYOLRE 600
OY 563 LAKFLHMLMSYVVELLSFEYVETTFQKNRLEFYRKSVMKLSIGIRQHLKRYOLRE 622
Db 601 LSEAEVROHREARPALTLTSLRLEFKPDGLRPIVMNDYVVGARTFRREKRAERLTSRYKA 660
OY 623 LSEAEVROHREARPALTLTSLRLEFKPDGLRPIVMNDYVVGARTFRREKRAERLTSRYKA 682
Db 661 LFSVYINTERARPPGLLGSVYLGLDIDIRAMRTFVLRYAODPPPELYTKVDYTAADTI 720
OY 683 LFSVYINTERARPPGLLGSVYLGLDIDIRAMRTFVLRYAODPPPELYTKVDYTAADTI 742
Db 721 PODRLTEVIASIIKPONTYCVRYAVYOKAAGHVRKAFAKSHVSTLTDLOPYMRQFVAHL 780
OY 743 PODRLTEVIASIIKPONTYCVRYAVYOKAAGHVRKAFAKSHVSTLTDLOPYMRQFVAHL 802
Db 781 QETSPLRDAVYIEQSSSLNEASSGLFDVFLRFMCHHAVIRKSGSYVOCGIPQSGIISLTL 840
OY 803 QETSPLRDAVYIEQSSSLNEASSGLFDVFLRFMCHHAVIRKSGSYVOCGIPQSGIISLTL 862
Db 841 LCSLCYGMENKLFAGIRDDGLLRVLVDLFTVPHLTHAKFTLTVRANGVEYICVYNL 900
OY 863 LCSLCYGMENKLFAGIRDDGLLRVLVDLFTVPHLTHAKFTLTVRANGVEYICVYNL 922
Db 901 RKTIVNPFVEDEALGTAFAVQMPAHGLFPWCGILLDTRTLEQSYDSYAFRISASLTF 960
OY 923 RKTIVNPFVEDEALGTAFAVQMPAHGLFPWCGILLDTRTLEQSYDSYAFRISASLTF 982
Db 961 NRGFAGRNMRKLFGLVRLKCHSLFLDLQVNSLQTVCTNIYKILLQAYRPHACVLOLP 1020
OY 983 NRGFAGRNMRKLFGLVRLKCHSLFLDLQVNSLQTVCTNIYKILLQAYRPHACVLOLP 1042
Db 1021 FHOQWKNKPTFFLRYISDTASLCYSILKAKNAGMSLGAKGAGPLPSEAVYQMLCHQAFLL 1080
OY 1043 FHOQWKNKPTFFLRYISDTASLCYSILKAKNAGMSLGAKGAGPLPSEAVYQMLCHQAFLL 1102
Db 1081 KLTRHRTVYVPLGLSLRTAQOTLSRKLPTCTTJALEAANPALPSDFKTIID 1132
OY 1103 KLTRHRTVYVPLGLSLRTAQOTLSRKLPTCTTJALEAANPALPSDFKTIID 1154

```

```

RA MEYERSON M., COUNTER C.M., EATON E.N., ELLISEN L.W., STEINER P.,
RA CADDE S.D., ZINIGRA L., BEIJERSBERGEN R.L., DAVIDOFF M.J., LIU Q.,
RA BACCHETTI S., HABER D.A., WEINBERG R.A.,
RL CELL 90:785-795(1997),
DR EMBL: AF018167; G2347129;
SQ SEQUENCE 1132 AA; 126937 MW; C1E5E2AF CRC32;

Query Match      98.1%; Score 8459; DB 4; Length 1132;
Best Local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 1131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 MPAPRCRAVYSLIRSHYREVLPATEVRRLGPOGWRLYOGRDPAARALVAOCLVCPW 60
OY 23 MPAPRCRAVYSLIRSHYREVLPATEVRRLGPOGWRLYOGRDPAARALVAOCLVCPW 82
Db 61 DARPPAPASTROYSCLELVARVYLQRLCEGAKNVYAFGALLDGRGPPPEAFTTSYR 120
OY 83 DARPPAPASTROYSCLELVARVYLQRLCEGAKNVYAFGALLDGRGPPPEAFTTSYR 142
Db 121 SYLPNTVTDALRGSGANGILLRRYGGDVLYVHLARCALFYLVAPSCAYOYCGPPLYOLGA 180
OY 143 SYLPNTVTDALRGSGANGILLRRYGGDVLYVHLARCALFYLVAPSCAYOYCGPPLYOLGA 202
Db 181 ATOARPPPHASGPRRRLGCERAMNHSYREAGVPLGLPAPGARRRGGSASRSLPLKRPBR 240
OY 203 ATOARPPPHASGPRRRLGCERAMNHSYREAGVPLGLPAPGARRRGGSASRSLPLKRPBR 262
Db 241 GAAPERPTVPGGSMWHPGRTGSPDGRGFCVVSAPARPAEATSELEGALSGTRSHSPSVG 300
OY 263 GAAPERPTVPGGSMWHPGRTGSPDGRGFCVVSAPARPAEATSELEGALSGTRSHSPSVG 322
Db 301 RQNHAGPSTSPRPMPGTPCPRPVYAEKHFILYSSGDKQQLRPSFLSSLSRSLTGARL 360
OY 323 RQNHAGPSTSPRPMPGTPCPRPVYAEKHFILYSSGDKQQLRPSFLSSLSRSLTGARL 382
Db 361 VETIFLGRPMWPGTPRRLPRLPQRYWOMRPLFELLGNHACCPGYLKTCPRLAAVT 420
OY 383 VETIFLGRPMWPGTPRRLPRLPQRYWOMRPLFELLGNHACCPGYLKTCPRLAAVT 442
Db 421 PAAGCAREKPGQSVAAABEEDTDPRRLVQLLRQSSPMQYGVYFACLRRLVPPGLWS 480
OY 443 PAAGCAREKPGQSVAAABEEDTDPRRLVQLLRQSSPMQYGVYFACLRRLVPPGLWS 502
Db 461 RHNERFLRNTKFTISLQKHAHLSLQELTKMSYVDCAMLRSPGVGCYPAAEHLRBEI 540
OY 503 RHNERFLRNTKFTISLQKHAHLSLQELTKMSYVDCAMLRSPGVGCYPAAEHLRBEI 562
Db 541 LAKFLHMLMSYVVELLSFEYVETTFQKNRLEFYRKSVMKLSIGIRQHLKRYOLRE 600
OY 563 LAKFLHMLMSYVVELLSFEYVETTFQKNRLEFYRKSVMKLSIGIRQHLKRYOLRE 622
Db 601 LSEAEVROHREARPALTLTSLRLEFKPDGLRPIVMNDYVVGARTFRREKRAERLTSRYKA 660
OY 623 LSEAEVROHREARPALTLTSLRLEFKPDGLRPIVMNDYVVGARTFRREKRAERLTSRYKA 682
Db 661 LFSVYINTERARPPGLLGSVYLGLDIDIRAMRTFVLRYAODPPPELYTKVDYTAADTI 720
OY 683 LFSVYINTERARPPGLLGSVYLGLDIDIRAMRTFVLRYAODPPPELYTKVDYTAADTI 742
Db 721 PODRLTEVIASIIKPONTYCVRYAVYOKAAGHVRKAFAKSHVSTLTDLOPYMRQFVAHL 780
OY 743 PODRLTEVIASIIKPONTYCVRYAVYOKAAGHVRKAFAKSHVSTLTDLOPYMRQFVAHL 802
Db 781 QETSPLRDAVYIEQSSSLNEASSGLFDVFLRFMCHHAVIRKSGSYVOCGIPQSGIISLTL 840
OY 803 QETSPLRDAVYIEQSSSLNEASSGLFDVFLRFMCHHAVIRKSGSYVOCGIPQSGIISLTL 862
Db 841 LCSLCYGMENKLFAGIRDDGLLRVLVDLFTVPHLTHAKFTLTVRANGVEYICVYNL 900
OY 863 LCSLCYGMENKLFAGIRDDGLLRVLVDLFTVPHLTHAKFTLTVRANGVEYICVYNL 922
Db 901 RKTIVNPFVEDEALGTAFAVQMPAHGLFPWCGILLDTRTLEQSYDSYAFRISASLTF 960

```

```

QY 923 RTIVNPEVEDEALGCTAFVQPAHGLFPMCGLLDTRTLEVOSSDYASRTASIRASLTF 982
Db 961 NNGFKAGMRRRLRGLVLRKCHSLFDLOVNSLOTVCTNITKILLLOAYRPAQLOLP 1020
QY 983 NNGFKAGMRRRLRGLVLRKCHSLFDLOVNSLOTVCTNITKILLLOAYRPAQLOLP 1042
Db 1021 FHOQVKNPFFELRVISDTASLCYSILKAKNAGMSLGAGAGPLSEAVQWMLCHQAFLL 1080
QY 1043 FHOQVKNPFFELRVISDTASLCYSILKAKNAGMSLGAGAGPLSEAVQWMLCHQAFLL 1102
Db 1081 KLTRHRVTVPLGLSLRTAQOTLSRKLPGTTLTALBAANPALPSDFKTLTD 1132
QY 1103 KLTRHRVTVPLGLSLRTAQOTLSRKLPGTTLTALBAANPALPSDFKTLTD 1154

RESULT 3
ID 070372 PRELIMINARY: PRT: 1122 AA.
AC 070372;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE TELOMERASE REVERSE TRANSCRIPTASE.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
RN [1]
RP SEQUENCE FROM N.A.
RA GREENBERG R.A., ALLSOPP R.C., CHIN L., MORIN G.B., DEPINHO R.A.;
RL ONCOGENE 0:0-0(1998).
DR EMBL, AF051911; G3005592; -.
KW RNA-DIRECTED DNA POLYMERASE.
SQ SEQUENCE 1122 AA; 127977 MW; 22207566 CRC32;

Query Match 59.8%; Score 5161; DB 11; Length 1122;
Best Local Similarity 62.7%; Pred. No. 0.00e+00;
Matches 718; Conservative 198; Mismatches 192; Indels 38; Gaps 24;

Db 1 MRRAPRCVAVRSLRSGREYEWPLATFVRRLGEGRRLOVQGPDKYIRTYVACVCMH 60
QY 23 MRRAPRCVAVRSLRSGREYEWPLATFVRRLGEGRRLOVQGPDKYIRTYVACVCMH 82
Db 61 GSOPPRADLSFHQVSLKELVAVVOVORLCEERNEVNLAFGEFELNEARGGPPMAFTSSVR 120
QY 83 DARPPRAASFPQVSLKELVAVVOVORLCEERNEVNLAFGEFELNEARGGPPMAFTSSVR 142
Db 121 STLPNTVIEFLRVSGAMLLSRVGGDLVYLLAHACALYLVPPSCAYOVCSSPLYQICA 180
QY 143 STLPNTVIEFLRVSGAMLLSRVGGDLVYLLAHACALYLVPPSCAYOVCSSPLYQICA 202
Db 181 TTDIMSVASASIRPTPVGRNFTNLFLOQIKSSROQAPRLALSRGKRLSLTSTS 240
QY 203 ATQARPPRAAS-GPRRLGCE-RA-W--NH--SVR-EGVPLGLPAGARRRGSSASHS 253
Db 241 VPSAKKARCVPRVREGEHROVLPTRPSGKSWPSPARS-PEV-PT-AEKDLS-S-KGYS 296
QY 254 LPLKPRRGARPERPRTRVYOGGSMHPRGTRGPS--DRGCVSPARPPEATSLGSAIS 312
Db 297 DLSLS-GSYCKHKRPSSTLSLSPRONAOLAR-FLETIRHFLSRDGOERLNSPFLSLN 354
QY 313 GTRHSHPSVGRQHNAGRPSTSRPRPMDPCRPVUAEKTKFLVSSD-KEQLRPSFLSS 371
Db 355 LQPNLTGAARLYEITFLGSRPRTSGPLCRNHSRKRYQMWRPFLFOOLLVNHACQVRL 414
QY 372 LRPSLTGAARLYEITFLGSRPRTSGPLCRNHSRKRYQMWRPFLFOOLLVNHACQVRL 431
Db 415 RSHCRPR---T-AN-----O--O--VT--DALUTSPHLMDLRLSSPWQVYGFRL 459
QY 432 KTHCPRLAAVTPAAGVCAHEKPGSVAAPEEDTDPRRLVQLRKHSSPWQVYGFRL 491
Db 460 CKVVSASLWGTNRNRRERFKNLKFTSLGKYGRLSLOELMKMKKVEDCHWLRSPPKDRY 519
QY 492 RRLVPPGLMGSRRNERRFLRNTKTKFTSLGKHAKLSTLOELWKKSVDCAMLRSPGVGC 551

```

```

Db 520 PAEHRRLREILATEFLWIMDYVOLLRSFFYITESTPOKRLRFFYRKSVMSKLSIGV 579
QY 552 PAEHRRLREILAKFLHMLMSYVVELLSFFYETTPQKRLRFFYRKSVMSKLSIGI 611
Db 580 ROHLERVLRELISOEVNRHODTWLAMPICRLFTPKPGLRPLVNMYSMSKTRALGRK 639
QY 612 ROHLKRVQLRELSEAEVRORHREARALLTSRLRFTPKPGLRPLVNMVYVGARFRERK 671
Db 640 QAOHFQRLKTFSLMNTERTKHPHLMGSSVYGMNDIYNTWAFYLRVALLOTPTMYTV 699
QY 672 RAERLTSRVKALFVLYNERARRPGLLGASVGLDIDHRAMTFVLYRVAADPPPLFV 731
Db 700 KADVTGAYADIPQKLVENVANMIRSESTYCIRQAVVRRSOGCVHKSFPROYTLTD 759
QY 732 KYDVYGAIDTIPQDRLEVIASITK-PQNTYCVRAVYQKAAHGVRAKAFKSHVSTLTD 790
Db 760 LQPYMGQFLKHLQDSDASALRNSVYIEOSISNMSSSSLFDEFLEHRLHSVYKIGDRCT 819
QY 791 LQPYMGQFAVHLQET--SPLRDVAVYIEQSSSLNEASSGLDFVFLRFMCHHAVRIRGKSYV 848
Db 820 QCOGIPQGSSTLTLCSLCFCGEMENKLPRAVQRODGLLRFVDEFLVPHLQAKTFLST 879
QY 849 QCOGIPQGSSTLTLCSLCFCGEMENKLPRAVQRODGLLRFVDEFLVPHLQAKTFLST 908
Db 880 LVHGVPEYGCMTLQKTVYVNEPVEPTLGAAPVYOLPAHCPFLMCGLLDTRTLEVEFCDY 939
QY 909 LVHGVPEYGCMTLQKTVYVNEPVEPTLGAAPVYOLPAHCPFLMCGLLDTRTLEVEFCDY 968
Db 940 SGYAOISITSLTFOSVFRAGTMRKLLSVLRCHGLFLDLQVNSLOTVCTNITKIFL 999
QY 969 SGYARTSIRASLTFNGFAGRMRRKLEGLVLRKCHSLFDLOVNSLOTVCTNITKILL 1028
Db 1000 LQAYRFAVYQLPFDQARKULTEFLGISSQASCCVAILKKNPGRML-K-ASGSP 1056
QY 1029 LQAYRFAVYQLPFDQARKULTEFLGISSQASCCVAILKKNPGRML-K-ASGSP 1088
Db 1057 PEAHMLCYQAFLLKLAHSVLYKCLGLPRTAOKLCKRLPEATMTILKAADALPSLD 1116
QY 1089 SEAVQMLCHQAFLLKTRHRVTVPLGLSLRTAQOTLSRKLPGTTLTALBAANPALPSD 1148
Db 1117 FQTLTD 1122
QY 1149 FQTLTD 1154

RESULT 4
ID 013339 PRELIMINARY: PRT: 988 AA.
AC 013339;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE TELOMERASE REVERSE TRANSCRIPTASE 1.
GN TR1.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=972H-.
RC MEDLINE: 97400623.
RX NAKAMURA T.M., MORIN G.B., CHAPMAN K.B., WEINRICH S.L., ANDREWS W.H.,
RA LINGNER J., HARLEY C.B., CECI T.R.;
RL SCIENCE 277:955-959(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-.
RA LYNE M., RAJANDREAM M.A., BARRELL B.G., VOLCKAERT G.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL, AF015783; G2340168; -.
DR PFAM; PF00078; tvt.
KW RNA-DIRECTED DNA POLYMERASE.
SQ SEQUENCE 988 AA; 116328 MW; 39C385A7 CRC32;

```



Db 651 FRKEMKDFROKFOKIALEGGOPYTTFLENEONDLNAKTLIVEAKORNYFEKNML 710  
QY 773 AHGVRRA-FKSHSTLT-DIOPYKOF-V-AHQETSPLDVAIVESSSINEMSSLE 828  
Db 711 QPVIN-ICQYVINEKNGKQYKQKIGIPGCLVSSILSSFYATLEESSIGELRDESNME 769  
QY 829 DVFLRFMCHN-VIRKRSYVOCGIPGSLSTLSCIGDMENK-L-F---AGIRD 882  
Db 770 NPNVNLRLDDVLLITTOENNAVLFTEKLINSRNGEFPFNMAKLOTSPPLSPSKFAK 829  
QY 883 G-----LLRLVDDELVLPHLTHATFRTLVGVPEYGVNFKTYVNFVEDEALG- 937  
Db 830 YGMDSEONIVQDCMDIGSIDAKTLAMPNIN-LIEGICLTINMQTKKASMLK 888  
QY 938 -GFAVQMPAHGL-F-PWGLDTRLEVOSSVARSSTIRASTN-KGFKGRNR 993  
Db 889 KKLKSL-MNNITVFRKTTTDEPANKTLNKLFISSGYKYMOCAKE--YKDFEKNLDM 945  
QY 994 KKLGVLLKCHSLFLDQVNSLQVCTINIKYKILLQVRRHACVLOLPHQOVWKNPTE 1053  
Db 946 SSMIDLEVSKITTYSTRA 963  
QY 1054 FLRVISDTASLCYSILKA 1071

RESULT 7  
ID 035432 PRELIMINARY; PRT: 67 AA.  
AC 035432;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE TELOMERASE CATALYTIC SUBUNIT (FRAGMENT).  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAQOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA DRIEST R., CLEVELAND J.L.;  
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; AF029235; G2605903; -;  
FT NON TER 1 1  
SQ SEQUENCE 67 AA; 8368 MM; E2A06F2B CRC32;  
SO

Query Match 4.9%; Score 424; DB 11; Length 67;  
Best Local Similarity 76.1%; Pred. No. 2,256-57;  
Matches 51; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

Db 1 FFYTESFQNRRLFFYKSVWSKIOSIGVROHLERVLRLSEVFAHHODTWLAMPIC 60  
QY 582 FFYETTFQNRRLFFYKSVWSKIOSIGVROHLERVLRLSEVFAHHODTWLAMPIC 641  
Db 61 RLRFIPK 67  
QY 642 RLRFIPK 648

RESULT 8  
ID 006163 PRELIMINARY; PRT: 884 AA.  
AC 006163;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE CHROMOSOME XII COSMID 8543.  
GN L8543.12.  
OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-S288C (AB972);  
RX MEDLINE; 97313267.  
RA JOHNSTON M., HILLIER L., RILES L., ALBERMANN K., ANDRE B., ANSGORGE W.,

RA BENES V., BRUCKNER M., DELIUS H., DUBOIS E., DUSTERHOFF A.,  
RA ENTIAN K.D., FLOETH M., GOFFEAU A., HEHLING U., HEWMAN K.,  
RA HENSS-NEITZEL D., HILBERT H., HILGER F., KLEINE K., KOTTER P.,  
RA LOUIS E.J., MESSENGUY F., MEMES H.W., MOSCA T., MOSTL D.,  
RA MOLLER-AUER S., NENTWICH U., OBERMAIER B., PIRAVIDI E., POHL T.M.,  
RA POTTELELE D., PURNELLE B., RECHMANN S., RIEGER M., RINKE M.,  
RA SCHARRE M., SCHERER B., SCHOLLER P., SCHWAGER C., SCHWARTZ S.,  
RA UNDERWOOD A.P., UNRESTRAU L.A., VANDENBOU M., VERHAASELT P.,  
RA VIERENDELS F., VOLT M., VOLKAERT G., VOSS H., WAMBITT R., WEDLER E.,  
RA WEDLER H., ZIMMERMAN F.K., ZOLLNER A., HANI J., HOEBISL J.D.;  
RL NATURE 387:0-0(0).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C (AB972);  
RA DU 2.1.  
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C (AB972);  
RA WATERSTON R.;  
RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C (AB972);  
RA CHERRY J.M.;  
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; U20618; G662136; -;  
SQ SEQUENCE 884 AA; 102662 MM; 1A94320F CRC32;  
SO

Query Match 3.6%; Score 314; DB 3; Length 884;  
Best Local Similarity 24.3%; Pred. No. 1,716-35;  
Matches 118; Conservative 122; Mismatches 202; Indels 44; Gaps 37;

Db 271 LSHSROSPKER-VLKFIYVILQILQEMFGSKKNGKIINKNLISLPNGLYPDS 329  
QY 470 LVQLRQSSPMQYGFVACLRRLVPGGLWGSRRNERFRNTKFKSLGKHAKSIOE 529  
Db 330 LKRLKLDKDFWLFIS-DIWTKNFENTNOLAIC-FISWFLQRLIPRIIOTFYCFIS 387  
QY 530 LTKKSVKDCAMLRSPGVCPAERHRLRELLAKFLHMLMSYVVELLSFFYVTEIT 589  
Db 388 STVTIYVE-RHDTWNKLTTPRIYEXFKTY-LVE-NNV-CRHNSYTSNENSHKRIIPK 443  
QY 590 FQKNRLFFYKSVWSKIOSIGVROHLERVLRLSEVFAHHODTWLAMPIC 648  
Db 444 KSNNEFRILAPCRGADDEEFTIYKEMK-NA-IQPTOKILEYLRNKRPTSF-TKISPT 500  
QY 649 -PDGLRPIVNNDY-VGARFTRERKRAERLTSRYKALFSLVNTYERARRPGLGASVGLD 706  
Db 501 QIADRT-KEFKQRLKKNFNVLPILFKFVYKSCYDIPMEQMR-ILKDALNENGFF 558  
QY 707 DI-HRAWTFVLRY-RA-QDPPPELYFKVDTGAYDIPQ-DRLTEVIASIIKPOMYC 762  
Db 559 VRSOY-FENTNT-G-VLKLFNV-VNASRVPKY-ELYDYNR-VVHLSNOVIN--VV- 608  
QY 763 VR-RVAVQKAAHGHVRAKRSVSTLTDLPYKQFAHLOETSPLDVAIVIQSSSLN 821  
Db 609 E---M-EIF-K-T---ALWEDKCYIREDELFOGSSISAPIVLDLVYDDLEFYSEFKAS 658  
QY 822 EASSGLDVLRFMCHNAVRIRKGSYVOCGIPGSLSTLSCIGDM-E-KKL-F-AG 878  
Db 659 PSQTLIKLADFLIISTDOQOYINIKKLAMGFGQYKNAKAN-RDKTILAVSSODDVT 717  
QY 879 IRRQGLRLVDLVLVYPLHILTHAKTFLKTLVGVPEYGVNFKTYVNFVEDEALG 938  
Db 718 IQFCAM 723  
QY 939 TAFVOM 944

RESULT 9  
ID 069118 PRELIMINARY; PRT: 296 AA.  
AC 069118;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE HYPOTHETICAL PROTEIN (FRAGMENT).  
 OS HUMAN HERPESVIRUS TYPE 4.  
 OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE;  
 OC GAMMAHERPESVIRINAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 87284169.  
 RA PEITZNER A.J., TSAI E.C., STROMINGER J.L., SPECK S.H.;  
 RL J. VIROL. 61:2902-2909(1987).  
 DR EMBL; M17294; G807646; -.  
 KM HYPOTHETICAL PROTEIN.  
 FT NON\_TER  
 SQ SEQUENCE 296 AA; 31393 MW; D17CE6F CRC32;  
 Query Match 1.7%; Score 147; DB 14; Length 296;  
 Best Local Similarity 38.6%; Pred. No. 7.18e-06;  
 Matches 39; Conservative 20; Mismatches 33; Indels 9; Gaps 8;  
 Db 91 GPPPTSGAAQ-RTHRRPCCPSANPCCPTWRR--R-SGAQKHPPPGAGQPSGP 146  
 Oy 194 GPPYQLGAATQARPAPHASG-PRR-R-IGCERAMNHSVREAGVPLGAPGA-RRRGS 249  
 Db 147 TGGRRAPGAPGTPAARGPGGAAVPSGATPHRSGSPAD 187  
 Oy 250 ASRSLPLKRRRGAAPER-TEVGQGSMAHPGRTGSPD 289  
 RESULT 10  
 ID 036027; PRELIMINARY; PRT; 574 AA.  
 AC 036027;  
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
 DE WISKOTT-ALDRICH SYNDROME PROTEIN HOMOLOG 1.  
 OS WSP1 OR SPAC4F10.15C  
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JS21;  
 RA ZANKEL T.C., OW D.W.;  
 RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-972;  
 RA CONNOR R., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WOOD V.;  
 RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- SIMILARITY: TO YEAST LAS17.  
 DR EMBL; AF038575; G2708709; -.  
 FT DOMAIN 311 317 POLY-PRO.  
 FT DOMAIN 337 343 POLY-PRO.  
 FT DOMAIN 361 366 POLY-PRO.  
 FT VARIANT 248 248 L -> V (IN STRAIN JS21).  
 SQ SEQUENCE 574 AA; 59605 MW; 9E84D37C CRC32;  
 Query Match 1.7%; Score 143; DB 3; Length 574;  
 Best Local Similarity 28.8%; Pred. No. 2.85e-05;  
 Matches 42; Conservative 41; Mismatches 53; Indels 10; Gaps 7;  
 Db 305 AANKKRPSPR-PSRR--NRG-KPIGNSNSLPPPPPSNAAG-SIPLPQGR 358  
 Oy 202 AATGARPAPASGRRLRGGERAMNHSVREAGVPLGAPGARRGSGASLPLPRR 261  
 Db 359 SAPPSPRSAPSTGROAPPLISSRAVSNP--APPALPGRSAPALPLGNASTSTPP 416  
 Oy 262 RGAAPDEPRTPVGGSWAHP-GRTGSPSDRGFCVSPARPAEATSLGALSGTRSHSPS 320  
 Db 417 VPTPP-SLPPSAPSLPPSAPSLPM 441

Oy 321 VGRQHNAGPSTSRPRMDTCCPV 346  
 RESULT 11  
 ID 000600; PRELIMINARY; PRT; 234 AA.  
 AC 000600;  
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
 DE PAROTID 'O' PROTEIN (FRAGMENT).  
 OS PRB4.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 96108975.  
 RA AZEN E.A., AMBERGER E., FISHER S., PRAKOBPHOL A., NIEGE R.L.;  
 RL AM. J. HUM. GENET. 58:143-153(1996).  
 DR EMBL; S80916; G1911492; -.  
 FT NON\_TER  
 SQ SEQUENCE 234 AA; 23656 MW; B5920075 CRC32;  
 Query Match 1.6%; Score 142; DB 4; Length 234;  
 Best Local Similarity 28.2%; Pred. No. 4.02e-05;  
 Matches 40; Conservative 41; Mismatches 52; Indels 9; Gaps 9;  
 Db 92 QGNGSOG-PPPHGRERPP-OGG-NQSHRPPPP-GKPER-PPPGGNGSOG-P-PP 144  
 Oy 199 QGAAATQARPAPASGRRLRGGERAMNHSVREAGVPLGAPGARRGSGASLPLPK 258  
 Db 145 HPGKPEPPPOEKNKRSASPPGKQPGQDGNPPGPPGKQPGPPGNGNQOQ 204  
 Oy 259 RPRGAAPERERTPVGGSWAHPGRTGSPD-R-GFCVSPAPAE-EAUSLEGALSGTHK 316  
 Db 205 APPAGKPGPPPPGGRPPR 226  
 Oy 317 SHPSVGRQHNAGPSTSRPRP 338  
 RESULT 12  
 ID 046612; PRELIMINARY; PRT; 276 AA.  
 AC 046612;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DE IS 1222 GENE ORF-A AND ORF-B.  
 OS ENTEROBACTER AGGLOMERANS.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 OC ENTEROBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-KLEBERGER, 1983;  
 RX MEDLINE; 9525664.  
 RA STEIBL H.D., LEMCKE F.M.;  
 RL GENE 156:37-42(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA STEIBL H.D., SIDDAVATTAM D.;  
 RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 96422758.  
 RA STEIBL H.D., SIDDAVATTAM D., KLINGMUELLER W.;  
 RL PLASMID 34:223-228(1995).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA STEIBL H.D.;  
 RL THESES (1995), UNIVERSITAET BAYREUTH.  
 DR EMBL; X78052; G459248; -.  
 DR EMBL; X81893; E258949; -.  
 SQ SEQUENCE 276 AA; 31718 MW; 0B54A420 CRC32;





\*\*\*\*\*  
MPSrch - protein database search, using Smith-Waterman algorithm  
Run on: Fri Dec 18 18:30:31 1998; MasPar time 9.93 Seconds  
Tabular output not generated. 823.071 Million cell updates/sec  
\*\*\*\*\*

Release 3.1a John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

\*\*\*\*\*  
Title: >US-08-951-733-20  
Description: (1-1154) from US08951733.pep  
Perfect Score: 8624  
Sequence: 1 HASGQRCVLRTEALAPAT.....TALEAANPALPSDFETILD 1154  
Scoring table: PAM 150  
Gap 11  
Searched: 77309 seqs, 7078906 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: a-issued  
1:5\_COMB 2:PCT9\_COMB 3:backfilest1  
Statistics: Mean 37.265; Variance 182.038; scale 0.205  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.  
SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	118	1.4	399	1	US-08-147-Sequence 27, Applicati	2.06e+00
2	118	1.4	399	1	US-08-447-Sequence 27, Applicati	2.06e+00
3	118	1.4	399	2	PCT-US92-0-Sequence 23, Applicati	2.06e+00
4	118	1.4	399	1	US-08-278-Sequence 23, Applicati	2.06e+00
5	118	1.4	399	1	US-08-643-Sequence 23, Applicati	2.06e+00
6	118	1.4	399	1	US-07-901-Sequence 13, Applicati	2.06e+00
7	118	1.4	399	1	US-08-643-Sequence 13, Applicati	2.06e+00
8	118	1.4	399	1	US-08-479-Sequence 10, Applicati	2.06e+00
9	118	1.4	399	1	US-08-155-Sequence 23, Applicati	2.06e+00
10	118	1.4	399	2	PCT-US93-0-Sequence 23, Applicati	2.06e+00
11	118	1.4	399	2	PCT-US93-1-Sequence 10, Applicati	2.06e+00
12	118	1.4	399	2	PCT-US93-0-Sequence 23, Applicati	2.06e+00
13	118	1.4	399	2	PCT-US93-0-Sequence 23, Applicati	2.06e+00
14	118	1.4	399	2	PCT-US93-0-Sequence 23, Applicati	2.06e+00
15	118	1.4	399	2	PCT-US93-0-Sequence 23, Applicati	2.06e+00
16	118	1.4	399	2	PCT-US93-0-Sequence 23, Applicati	2.06e+00
17	118	1.4	399	1	US-08-480-Sequence 10, Applicati	2.06e+00
18	118	1.4	399	2	PCT-US93-0-Sequence 23, Applicati	2.06e+00
19	118	1.4	399	1	US-08-406-Sequence 23, Applicati	2.06e+00
20	118	1.4	399	1	US-08-451-Sequence 23, Applicati	2.06e+00
21	118	1.4	399	2	PCT-US93-0-Sequence 23, Applicati	2.06e+00
22	118	1.4	399	2	PCT-US91-0-Sequence 6, Applicati	2.06e+00
23	125	1.4	402	1	US-07-841-Sequence 29, Applicati	7.05e-01

24	125	1.4	402	1	US-08-643-Sequence 21, Applicati	7.05e-01
25	125	1.4	402	1	US-08-451-Sequence 21, Applicati	7.05e-01
26	125	1.4	402	1	US-08-462-Sequence 21, Applicati	7.05e-01
27	125	1.4	402	2	PCT-US93-0-Sequence 21, Applicati	7.05e-01
28	125	1.4	402	1	US-08-206-Sequence 4, Applicati	7.05e-01
29	125	1.4	402	1	US-08-155-Sequence 21, Applicati	7.05e-01
30	125	1.4	402	2	PCT-US93-0-Sequence 21, Applicati	7.05e-01
31	125	1.4	402	1	US-08-480-Sequence 8, Applicati	7.05e-01
32	125	1.4	402	1	US-07-901-Sequence 11, Applicati	7.05e-01
33	125	1.4	402	1	US-08-147-Sequence 29, Applicati	7.05e-01
34	125	1.4	402	1	US-08-643-Sequence 21, Applicati	7.05e-01
35	125	1.4	402	1	US-08-406-Sequence 8, Applicati	7.05e-01
36	125	1.4	402	1	US-08-479-Sequence 21, Applicati	7.05e-01
37	125	1.4	402	1	US-08-278-Sequence 21, Applicati	7.05e-01
38	125	1.4	402	1	US-08-447-Sequence 29, Applicati	7.05e-01
39	125	1.4	402	2	PCT-US93-0-Sequence 21, Applicati	7.05e-01
40	125	1.4	402	2	PCT-US93-0-Sequence 21, Applicati	7.05e-01
41	125	1.4	402	2	PCT-US93-0-Sequence 21, Applicati	7.05e-01
42	125	1.4	402	2	PCT-US93-1-Sequence 8, Applicati	7.05e-01
43	125	1.4	402	2	PCT-US92-0-Sequence 21, Applicati	7.05e-01
44	125	1.4	402	2	PCT-US93-0-Sequence 11, Applicati	7.05e-01
45	125	1.4	402	2	PCT-US93-0-Sequence 6, Applicati	7.05e-01

## ALIGNMENTS

RESULT 1  
ID US-08-147-023-27 STANDARD: PRT; 399 AA.  
XX  
AC  
XX  
Sequence 27, Application US/08147023  
DE Patent No. 5468845  
CC  
CC GENERAL INFORMATION:  
CC APPLICANT: OPPERMAN, HERMANN  
CC APPLICANT: OZKAYNAK, ENGIN  
CC APPLICANT: KUBERASAMPATH, THANGAVEL  
CC APPLICANT: RUEGER, DAVID C.  
CC APPLICANT: PANG, ROY H.L.  
CC TITLE OF INVENTION: OSTEOGENIC DEVICES  
CC NUMBER OF SEQUENCES: 33  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
CC STREET: 53 STATE STREET  
CC CITY: BOSTON  
CC STATE: MASSACHUSETTS  
CC COUNTRY: U.S.A.  
CC ZIP: 02109  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: IBM PC compatible  
CC SOFTWARE: Patent Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/147,023  
CC FILING DATE: 21-FEB-1992  
CC CLASSIFICATION: 530  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 810,560  
CC FILING DATE: 20-DEC-1991  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 827,052  
CC FILING DATE: 28-JAN-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 660,162  
CC FILING DATE: 22-FEB-1991  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 621,988  
CC FILING DATE: 04-DEC-1990

CC	Prior Application Data:	US 621, 849
CC	Application Number:	US 621, 849
CC	Filing Date:	04-DEC-1990
CC	Prior Application Data:	
CC	Application Number:	US 616, 374
CC	Filing Date:	21-NOV-1990
CC	Prior Application Data:	
CC	Application Number:	US 600, 024
CC	Filing Date:	18-OCT-1990
CC	Prior Application Data:	
CC	Application Number:	US 599, 543
CC	Filing Date:	18-OCT-1990
CC	Prior Application Data:	
CC	Application Number:	US 579, 865
CC	Filing Date:	07-SEP-1990
CC	Prior Application Data:	
CC	Application Number:	US 569, 920
CC	Filing Date:	20-AUG-1990
CC	Prior Application Data:	
CC	Application Number:	US 483, 913
CC	Filing Date:	22-FEB-1990
CC	Prior Application Data:	
CC	Application Number:	US 422, 613
CC	Filing Date:	17-OCT-1989
CC	Prior Application Data:	
CC	Application Number:	US 315, 342
CC	Filing Date:	23-FEB-1989
CC	Prior Application Data:	
CC	Application Number:	US 232, 630
CC	Filing Date:	15-AUG-1988
CC	Prior Application Data:	
CC	Application Number:	US 179, 460
CC	Filing Date:	08-APR-1988
CC	Attorney/Agent Information:	
CC	Name:	PITCHER, EDMUND R.
CC	Registration Number:	27, 829
CC	Reference/Docket Number:	CRP-001CP6
CC	Telecommunication Information:	
CC	Telephone:	617/248-7000
CC	Telefax:	617/248-7100
CC	Information for SEQ ID NO:	27
CC	Sequence Characteristics:	
CC	Length:	399 amino acids
CC	Type:	amino acid
CC	Topology:	linear
CC	Molecule Type:	protein
SQ	Sequence	399 AA; 44764 MW; 790568 CN;
Db	Query Match	1.4%; Score 118; DB 1; Length 399;
	Best Local Similarity	38.7%; Pred. NO. 2.06e+00;
QY	Matches 24; Conservative	12; Mismatches 22; Indels 4; Gaps 4;
12	GLALCALGGGSGPRPP-ITC-PQRRLG-ARRRDMQRRITLAVLGLPGR-PPRPAQPPAAR	67
194	GPLYQLGAATQARPPHASPGRRLRGCEAMNHSVREAGVPLGLPAPGARRRGSSASRS	253
Db	68 QP 69	
QY	254 LP 255	
RESULT	2	
ID	US-08-447-570-27	STANDARD; PRT; 399 AA.
AC	xxxxxx	
DT		
XX		
XX	Sequence 27, Application US/08447570	
DE		
XX	Sequence 27, Application US/08447570	
CC	Patent No. 5714589	
CC	GENERAL INFORMATION:	
CC		

```

CC APPLICANT: OPPERMAN, HERMAN
CC APPLICANT: OKRAYNA, ENGIN
CC APPLICANT: KUBERASAMPATH, THANNGAVEL
CC APPLICANT: ROEGER, DAVID C.
CC APPLICANT: PANG, ROY H.L.
CC TITLE OF INVENTION: OSTEOGENIC DEVICES
CC NUMBER OF SEQUENCES: 33
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: TESTA, HORWITZ & THIBAULT
CC STREET: 53 STATE STREET
CC CITY: BOSTON
CC STATE: MASSACHUSETTS
CC COUNTRY: U.S.A.
CC ZIP: 02109
CC
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/447,570
CC FILING DATE: 21-FEB-1992
CC CLASSIFICATION: 536
CC
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 810,560
CC FILING DATE: 20-DEC-1991
CC
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 827,052
CC FILING DATE: 28-JAN-1992
CC
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 660,162
CC FILING DATE: 22-FEB-1991
CC
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 621,988
CC FILING DATE: 04-DEC-1990
CC
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 621,849
CC FILING DATE: 04-DEC-1990
CC
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 616,374
CC FILING DATE: 21-NOV-1990
CC
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 600,024
CC FILING DATE: 18-OCT-1990
CC
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 599,543
CC FILING DATE: 18-OCT-1990
CC
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 579,865
CC FILING DATE: 07-SEP-1990
CC
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 569,920
CC FILING DATE: 20-AUG-1990
CC
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 483,913
CC FILING DATE: 22-FEB-1990
CC
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 422,613
CC FILING DATE: 17-OCT-1989
CC
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 315,342
CC FILING DATE: 23-FEB-1989
CC
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 232,630
CC FILING DATE: 15-AUG-1988
CC
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 179,460
CC FILING DATE: 08-APR-1988
CC
CC ATTORNEY/AGENT INFORMATION:
CC NAME: PITCHER, EDMUND R.
CC REGISTRATION NUMBER: 27,829
CC REFERENCE/DOCKET NUMBER: CRP-001CP6
CC TELECOMMUNICATION INFORMATION:

```

```

CC      TOPOLOGY: linear
CC      MOLECULE TYPE: protein
CC      SEQUENCE 399 AA; 44764 MW; 790568 CN;
SO
Query Match          1.4%; Score 118; DB 2; Length 399;
Best Local Similarity 38.7%; Pred. No. 2,06e+00;
Matches 24; Conservative 12; Mismatches 22; Indels 4; Gaps 4;
Db      12 GLATLCGGHGRPP-HTC-PQRRL-ARERRDMSREILAVIGLGR-PPRAQPAAR 67
      1 : 111 : 111 : 11111 : 111111 : 11111 : 11111 :
Oy      194 GPPIYQLGAAQTGAPPPHSGSPRRLLCERAMNHSVREAGVPLGLPAPGARRRGGSASRS 253
      1 : 111 : 111 : 11111 : 111111 : 11111 : 11111 :
Db      68 QP 69
      1 : 111 : 111 : 11111 : 111111 : 11111 : 11111 :
Oy      254 LP 255
      1 : 111 : 111 : 11111 : 111111 : 11111 : 11111 :
RESULT 4
ID      US-08-278-729A-23      STANDARD;      PRT;      399 AA.
XX      xxxxxx
XX      Sequence 23, Application US/08278729A
XX
CC      Sequence 23, Application US/08278729A
CC      Patent No. 5650276
CC      GENERAL INFORMATION:
CC      APPLICANT: SMART, JOHN
CC      APPLICANT: OPPERMAN, HERMAN
CC      APPLICANT: OZKATNAK, ENGIN
CC      APPLICANT: KUBERASAMPATH, THANGAVEL
CC      APPLICANT: RUEGER, DAVID C.
CC      APPLICANT: PANG, ROY H. L.
CC      APPLICANT: COHEN, CHARLES M.
CC      TITLE OF INVENTION: MORPHOGENIC PROTEIN SCREENING METHOD
CC      NUMBER OF SEQUENCES: 3
CC      CORRESPONDENCE ADDRESSES:
CC      ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
CC      STREET: 45 SOUTH STREET
CC      CITY: HOPKINTON
CC      STATE: MA
CC      COUNTRY: USA
CC      ZIP: 01748
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Floppy disk
CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: Patentin Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: US/08/278,729A
CC      FILING DATE: 20-JUL-1994
CC      CLASSIFICATION: 435
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: PITCHER Esq., EDMUND R.
CC      REGISTRATION NUMBER: 27,829
CC      REFERENCE/DOCKET NUMBER: CRP-058CPFW
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (508) 435-9001
CC      TELEFAX: (508) 435-6951
CC      INFORMATION FOR SEQ ID NO: 23:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 399 amino acids
CC      TYPE: amino acid
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: protein
CC      SEQUENCE 399 AA; 44764 MW; 790568 CN;
SO
Query Match          1.4%; Score 118; DB 1; Length 399;
Best Local Similarity 38.7%; Pred. No. 2,06e+00;
Matches 24; Conservative 12; Mismatches 22; Indels 4; Gaps 4;

```

Db	68	QP	69
QY	254	LP	255

DE	Sequence 23, Application US/08643563A
XX	
CC	Sequence 23, Application US/08643563A
CC	Patent No. 5707810



```

CC      MEDIUM TYPE: floppy disk
CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: Patent Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: US/08/155,343A
CC      FILING DATE: 15-NOV-1993
CC      CLASSIFICATION: 435
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: FENTON Esq., GILLIAN M.
CC      REGISTRATION NUMBER: 36,508
CC      REFERENCE/DOCKET NUMBER: CRP-067FW
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (617) 248-7560
CC      TELEFAX: (617) 248-7100
CC      INFORMATION FOR SEQ ID NO: 23:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 399 amino acids
CC      TYPE: amino acid
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: protein
CC      SEQUENCE 399 AA; 44764 MW; 790568 CN;
CC
CC      Query Match          1.4%; Score 118; DB 1; Length 399;
CC      Best Local Similarity 38.7%; Pred. NO. 2.06e+00;
CC      Matches 24; Conservative 12; Mismatches 22; Indels 4; Gaps 4.;
CC
Dd      12 GLALCALGGHGHPRP-HTC-PORRLG-ARBRDMOREILAVLGLPGR-PPRAQPAAR 67
Qy      194 GPPLVQLGAAVQARPPHASPFRRLRGERRAMWNVSVEAGVPLGLPAPGARRRGSGASRS 253
Dd      68 QP 69
Qy      254 LP 255
Qy
Qy      RESULT 10
ID      PCT-US93-07231-23 STANDARD; PRT: 399 AA.
XX      xxxxxx
XX
XX      Sequence 23, Application PC/TUS9307231
Dd
Dd      Sequence 23, Application PC/TUS9307231
XX
XX      GENERAL INFORMATION:
CC      APPLICANT:
CC      TITLE OF INVENTION: MORPHOGEN-INDUCED NERVE REGENERATION AND
CC      TITLE OF INVENTION: REPAIR
CC      NUMBER OF SEQUENCES: 33
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: CREATIVE BIOMOLECULES, INC.
CC      STREET: 35 SOUTH STREET
CC      CITY: HOPKINTON
CC      STATE: MASSACHUSETTS
CC      COUNTRY: USA
CC      ZIP: 01748
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: floppy disk
CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: Patent Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: PCT/US93/07231
CC      FILING DATE: 19930729
CC      CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: KELLEY, ROBIN D.
CC      REGISTRATION NUMBER: 34,637
CC      REFERENCE/DOCKET NUMBER: CRP-070
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: 617/248-7000
CC

```

```

CC      TELEFAX: 617/248-7100
CC      INFORMATION FOR SEQ. ID NO: 23:
CC      SEQUENCE CHARACTERISTICS:
CC          LENGTH: 399 amino acids
CC          TYPE: amino acid
CC          TOPOLOGY: linear
CC      MOLECULE TYPE: protein
CC      SEQUENCE 399 AA; 44764 MW; 790568 CN;
SQ
CC      Query Match 1.4%; Score 118; DB 2; Length 399;
CC      Best Local Similarity 38.7%; Pred. No. 2.06e+00;
CC      Matches 24; Conservative 12; Mismatches 22; Indels 4; Gaps 4;
Db
CC      12 GLAICAGGGGPPR-HTC-PORRIG-ANRRMDMOREILLAVIGLPER-PRPRAGPPAAR 67
CC      194 GPPLYQLGAAQAQAPPHPASGPRRRRLCERAMVNSVREAGVPLGLPAPGARRRGGSASRS 253
Cc
Cc      Db 68 QP 69
Cc      QY 254 LP 255
Cc
Cc      RESULT 11
Cc      ID PCT-US93-05446-13 STANDARD; PRT; 399 AA.
Cc      XX
Cc      XXXXXX
Cc
Cc      Sequence 13, Application PC/TUS9305446
Cc      DE
Cc      Sequence 13, Application PC/TUS9305446
Cc      CC GENERAL INFORMATION:
Cc      CC APPLICANT:
Cc      CC TITLE OF INVENTION: PROSTHETIC DEVICES HAVING ENHANCED
Cc      CC TITLE OF INVENTION: OSTEOGENIC PROPERTIES
Cc      CC NUMBER OF SEQUENCES: 22
Cc      CC CORRESPONDENCE ADDRESS:
Cc      CC ADDRESSEE: Creative Biomolecules, Inc.
Cc      CC STREET: 35 South Street
Cc      CC CITY: Hopkinton
Cc      CC STATE: MA
Cc      CC COUNTRY: USA
Cc      CC ZIP: 01748
Cc      CC COMPUTER READABLE FORM:
Cc      CC MEDIUM TYPE: Floppy disk
Cc      CC COMPUTER: IBM PC compatible
Cc      CC OPERATING SYSTEM: PC-DOS/MS-DOS
Cc      CC SOFTWARE: Patentin Release #1.0, Version #1.25
Cc      CC CURRENT APPLICATION DATA:
Cc      CC APPLICATION NUMBER: PCT/US93/05446
Cc      CC FILING DATE: 19930608
Cc      CC CLASSIFICATION:
Cc      CC ATTORNEY/AGENT INFORMATION:
Cc      CC NAME: PITCHER, ESQ., EDMUND R
Cc      CC REGISTRATION NUMBER: 27,829
Cc      CC REFERENCE/DOCKET NUMBER: STK-057
Cc      CC TELECOMMUNICATION INFORMATION:
Cc      CC TELEPHONE: 617/248-7000
Cc      CC INFORMATION FOR SEQ. ID NO: 13:
Cc      CC SEQUENCE CHARACTERISTICS:
Cc      CC LENGTH: 399 amino acids
Cc      CC TYPE: AMINO ACID
Cc      CC TOPOLOGY: linear
Cc      CC MOLECULE TYPE: protein
Cc      CC SEQUENCE 399 AA; 44764 MW; 790568 CN;
Cc
Cc      Query Match 1.4%; Score 118; DB 2; Length 399;
Cc      Best Local Similarity 38.7%; Pred. No. 2.06e+00;
Cc      Matches 24; Conservative 12; Mismatches 22; Indels 4; Gaps 4;
Cc      12 GLAICAGGGGPPR-HTC-PORRIG-ANRRMDMOREILLAVIGLPER-PRPRAGPPAAR 67
Cc      194 GPPLYQLGAAQAQAPPHPASGPRRRRLCERAMVNSVREAGVPLGLPAPGARRRGGSASRS 253

```

OY 194 GPPYLQAGATQARPPHASPBRRLGGERAMNHSVREAGVPLGLPAPGARRRGGSASRS 253  
DB 68 QP 69  
OY 254 LP 255

RESULT 12  
ID PCT-US93-10520-10 STANDARD: PRT: 399 AA.  
XX xxxxxx  
AC  
DT  
XX  
DE Sequence 10, Application PC/TUS9310520  
XX Sequence 10, Application PC/TUS9310520  
CC GENERAL INFORMATION:  
CC APPLICANT:  
CC TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS  
CC NUMBER OF SEQUENCES: 13  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: CREATIVE BIOMOLECULES, INC.  
CC STREET: 45 SOUTH STREET  
CC CITY: HOPKINTON  
CC STATE: MA  
CC COUNTRY: USA  
CC ZIP: 01748  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US93/10520  
CC FILING DATE:  
CC CLASSIFICATION:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: KELLEY, ROBIN D.  
CC REGISTRATION NUMBER: 34,637  
CC REFERENCE/DOCKET NUMBER: CRP-071  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 617/248-7000  
CC TELEFAX: 617/248-7100  
CC INFORMATION FOR SEQ ID NO: 23:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 399 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 399 AA: 44764 MW: 790568 CN;

Query Match 1.4%; Score 118; DB 2; Length 399;  
Best Local Similarity 38.7%; Pred. No. 2,06+00;  
Matches 24; Conservative 12; Mismatches 22; Indels 4; Gaps 4;

DB 12 GLALCALGGHGPPRP-HTC-PORRLG-ARERRDMOREITLAVLIGR-PPRPAOPAAAR 67  
OY 194 GPPYLQAGATQARPPHASPBRRLGGERAMNHSVREAGVPLGLPAPGARRRGGSASRS 253  
DB 68 QP 69  
OY 254 LP 255

RESULT 13  
ID PCT-US93-07190-23 STANDARD: PRT: 399 AA.  
XX xxxxxx  
AC  
DT  
XX  
DE Sequence 23, Application PC/TUS9307190  
XX Sequence 23, Application PC/TUS9307190  
CC GENERAL INFORMATION:  
CC APPLICANT:  
CC TITLE OF INVENTION: MORPHOGEN-ENRICHED DIETARY COMPOSITION  
CC NUMBER OF SEQUENCES: 33  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: CREATIVE BIOMOLECULES, INC.  
CC STREET: 35 SOUTH STREET  
CC CITY: HOPKINTON  
CC STATE: MA  
CC COUNTRY: USA  
CC ZIP: 01748  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US93/07190  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: KELLEY, ROBIN D.  
CC REGISTRATION NUMBER: 34,637  
CC REFERENCE/DOCKET NUMBER: CRP-071  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 617/248-7000  
CC TELEFAX: 617/248-7100  
CC INFORMATION FOR SEQ ID NO: 23:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 399 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 399 AA: 44764 MW: 790568 CN;

Query Match 1.4%; Score 118; DB 2; Length 399;  
Best Local Similarity 38.7%; Pred. No. 2,06+00;  
Matches 24; Conservative 12; Mismatches 22; Indels 4; Gaps 4;

DB 12 GLALCALGGHGPPRP-HTC-PORRLG-ARERRDMOREITLAVLIGR-PPRPAOPAAAR 67  
OY 194 GPPYLQAGATQARPPHASPBRRLGGERAMNHSVREAGVPLGLPAPGARRRGGSASRS 253  
DB 68 QP 69  
OY 254 LP 255

RESULT 14  
ID US-08-462-623-23 STANDARD: PRT: 399 AA.  
XX xxxxxx  
AC  
DT  
XX  
DE Sequence 23, Application US/08462623

XX Sequence 23, Application US/08462623  
CC Patent No. 5739107  
CC GENERAL INFORMATION:  
CC APPLICANT: COHEN, CHARLES M.  
CC APPLICANT: CHARETTE, MARC F.  
CC APPLICANT: KUBERASAMPATH, THANGAVEL  
CC APPLICANT: RUEGER, DAVID C.  
CC APPLICANT: OPPERMAN, HERMANN  
CC APPLICANT: PANG, ROY H.L.  
CC APPLICANT: OZKAYNAK, ENGIN  
CC APPLICANT: SMART, JOHN E.  
CC TITLE OF INVENTION: MORPHOGEN TREATMENT OF GASTROINTESTINAL  
CC TITLE OF INVENTION: ULCERS.  
CC NUMBER OF SEQUENCES: 33  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES  
CC STREET: 45 SOUTH STREET  
CC CITY: HOPKINTON  
CC STATE: MA  
CC COUNTRY: USA  
CC ZIP: 01748  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/462,623  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/445,862  
CC FILING DATE: 22-MAY-1995  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: FENTON Esq., GILLIAN M.  
CC REGISTRATION NUMBER: 36,508  
CC REFERENCE/DOCKET NUMBER: CRP-074CN  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (508) 435-9901  
CC TELEFAX: (508) 435-6951  
CC INFORMATION FOR SEQ ID NO: 23:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 399 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 399 AA; 44764 MW; 790568 CN;  
SQ  
Query Match 1.4%; Score 118; DB 1; Length 399;  
Best Local Similarity 38.7%; Pred. No. 2.06e+00;  
Matches 24; Conservative 12; Mismatches 22; Indels 4; Gaps 4;  
Db 12 GLAICALGGHGPRPP-HTC-PORRLG-ARERRDMOREILAVIGLGR-PRPRAQPAAR 67  
QY 194 GPPLYQLGAATQARPPPHASGPRRLGCEPAMNHSYREAGVPLGLPAPGARRRGSSASRS 253  
Db 68 QP 69  
QY 254 LP 255  
RESULT 15  
ID PCT-0593-07189-8 STANDARD; PRT; 399 AA.  
XX xxxxxx  
XX  
XX  
XX  
XX  
XX Sequence 8, Application PC/TUS9307189  
XX Sequence 8, Application PC/TUS9307189  
CC GENERAL INFORMATION:

CC APPLICANT:  
CC TITLE OF INVENTION: NOVEL MORPHOGENIC PROTEIN COMPOSITIONS  
CC TITLE OF INVENTION: OF MATTER  
CC NUMBER OF SEQUENCES: 23  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: PATENT ADMINISTRATOR/CREATIVE BIOMOLECULES,  
CC ADDRESSEE: INC.  
CC STREET: 35 SOUTH STREET  
CC CITY: HOPKINTON  
CC STATE: MA  
CC COUNTRY: USA  
CC ZIP: 01748  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/TUS93/07189  
CC FILING DATE: 19930729  
CC CLASSIFICATION:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER:  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: KELLEY, ROBIN D.  
CC REGISTRATION NUMBER: 34,637  
CC REFERENCE/DOCKET NUMBER: CRP-081CP  
CC INFORMATION FOR SEQ ID NO: 8:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 399 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 399 AA; 44764 MW; 790568 CN;  
SQ  
Query Match 1.4%; Score 118; DB 2; Length 399;  
Best Local Similarity 38.7%; Pred. No. 2.06e+00;  
Matches 24; Conservative 12; Mismatches 22; Indels 4; Gaps 4;  
Db 12 GLAICALGGHGPRPP-HTC-PORRLG-ARERRDMOREILAVIGLGR-PRPRAQPAAR 67  
QY 194 GPPLYQLGAATQARPPPHASGPRRLGCEPAMNHSYREAGVPLGLPAPGARRRGSSASRS 253  
Db 68 QP 69  
QY 254 LP 255  
Search completed: Fri Dec 18 18:31:03 1998  
Job time : 32 secs.

\*\*\*\*\*  
WIRE  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Dec 18 18:31:23 1998; MasPar time 55.03 Seconds

Tabular output not generated. 793.597 Million cell updates/sec

Title: >US-08-951-733-20

Description: (1-1154) from US08951733.pep

Perfect Score: 8624

Sequence: 1 HASGQRCVLTWELAPAT.....TALEAANPALPSDFITLD 1154

Scoring table: PAM 150

Searched: 321025 segs, 37841481 residues

Post-processing: Minimum Match 0%

Database: a-pending

1:P9 2:U60 3:U7 4:U80 5:U81 6:U82 7:U83 8:U84 9:U85

10:U86 11:U87 12:U88 13:U89 14:U90 15:U91 16:NEWP

17:NEWU6 18:NEWU8 19:NEWU9

Statistics: Mean 40.459; Variance 196.267; scale 0.206

## SUMMARIES

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	8624	100.0	1154	13	US-08-951-Sequence 20, Applicat	0.00e+00
2	8627	99.8	1189	13	US-08-974-Sequence 613, Applicat	0.00e+00
3	8607	99.8	1189	13	US-08-911-Sequence 34, Applicat	0.00e+00
4	8607	99.8	1189	13	US-08-912-Sequence 325, Applicat	0.00e+00
5	8607	99.8	1200	18	US-08-974-Sequence 612, Applicat	0.00e+00
6	8607	99.8	1200	13	US-08-912-Sequence 324, Applicat	0.00e+00
7	8607	99.8	1200	13	US-08-912-Sequence 33, Applicat	0.00e+00
8	8607	99.8	1285	18	US-08-874-Sequence 600, Applicat	0.00e+00
9	8607	99.8	1285	13	US-08-911-Sequence 32, Applicat	0.00e+00
10	8607	99.8	1285	13	US-08-912-Sequence 314, Applicat	0.00e+00
11	8590	99.6	1407	13	US-08-911-Sequence 55, Applicat	0.00e+00
12	8590	99.6	1407	18	US-08-974-Sequence 628, Applicat	0.00e+00
13	8500	99.6	1407	13	US-08-912-Sequence 334, Applicat	0.00e+00
14	8455	98.2	1132	18	US-08-974-Sequence 2, Applicatio	0.00e+00
15	8455	98.2	1132	13	US-08-912-Sequence 2, Applicatio	0.00e+00
16	8455	98.2	1132	13	US-08-911-Sequence 2, Applicatio	0.00e+00
17	8455	98.2	1132	14	US-09-042-Sequence 3, Applicatio	0.00e+00
18	8455	98.2	1132	12	US-09-052-Sequence 2, Applicatio	0.00e+00
19	8455	98.2	1132	14	US-08-854-Sequence 225, Applicat	0.00e+00
20	8455	98.2	1154	13	US-08-912-Sequence 322, Applicat	0.00e+00
21	8455	98.2	1154	18	US-08-974-Sequence 611, Applicat	0.00e+00

22	8465	98.2	1154	13	US-08-911-Sequence 45, Applicati	0.00e+00
23	8459	98.1	1132	14	US-09-026-Sequence 3, Applicatio	0.00e+00
24	8448	98.0	1132	18	US-08-974-Sequence 344, Applicat	0.00e+00
25	7113	82.5	949	13	US-08-951-Sequence 14, Applicati	0.00e+00
26	5742	66.6	807	13	US-08-912-Sequence 5, Applicatio	0.00e+00
27	5742	66.6	807	13	US-08-911-Sequence 19, Applicati	0.00e+00
28	5742	66.6	807	18	US-08-974-Sequence 5, Applicatio	0.00e+00
29	5334	61.9	1003	18	US-08-974-Sequence 336, Applicat	0.00e+00
30	5334	61.9	1003	12	US-08-854-Sequence 217, Applicat	0.00e+00
31	5334	61.9	1003	12	US-08-851-Sequence 217, Applicat	0.00e+00
32	5161	59.8	1122	14	US-09-042-Sequence 47, Applicatio	0.00e+00
33	4070	47.2	545	13	US-08-911-Sequence 267, Applicat	2.01e-199
34	2458	28.5	564	18	US-08-974-Sequence 101, Applicat	2.01e-199
35	2458	28.5	564	12	US-08-854-Sequence 101, Applicat	2.01e-199
36	2458	28.5	564	12	US-08-846-Sequence 101, Applicat	2.01e-199
37	2458	28.5	564	12	US-08-851-Sequence 101, Applicat	2.01e-199
38	2226	25.8	538	13	US-08-912-Sequence 316, Applicat	5.73e-179
39	2226	25.8	538	18	US-08-974-Sequence 602, Applicat	5.73e-179
40	2226	25.8	538	13	US-08-911-Sequence 46, Applicati	2.00e-170
41	2129	24.7	514	13	US-08-911-Sequence 41, Applicati	2.00e-170
42	2129	24.7	514	13	US-08-912-Sequence 319, Applicat	2.00e-170
43	2129	24.7	514	18	US-08-974-Sequence 603, Applicat	1.49e-166
44	2085	24.2	530	13	US-08-912-Sequence 317, Applicat	1.49e-166
45	2085	24.2	530	18	US-08-974-Sequence 603, Applicat	1.49e-166

ALIGNMENTS

RESULT 1	ID	US-08-951-733-20	STANDARD:	PRT;	1154 AA.
XX	XX	xxxxxx			
XX	XX				
DE	DE	Sequence 20, Application US/08951733			
CC	CC	Sequence 20, Application US/08951733			
CC	CC	GENERAL INFORMATION:			
CC	CC	APPLICANT: Harrington, Lea A.			
CC	CC	APPLICANT: Robinson, Murray O.			
CC	CC	TITLE OF INVENTION: NOVEL GENES ENCODING TELOMERASE PROTEINS			
CC	CC	NUMBER OF SEQUENCES: 44			
CC	CC	CORRESPONDENCE ADDRESS:			
CC	CC	ADDRESSEE: Amgen Inc.			
CC	CC	STREET: One Amgen Center Drive			
CC	CC	CITY: Thousand Oaks			
CC	CC	STATE: CA			
CC	CC	COUNTRY: USA			
CC	CC	ZIP: 91320-1789			
CC	CC	COMPUTER READABLE FORM:			
CC	CC	MEDIUM TYPE: Floppy disk			
CC	CC	COMPUTER: IBM PC compatible			
CC	CC	OPERATING SYSTEM: PC-DOS/MS-DOS			
CC	CC	SOFTWARE: Patentin Release #1.0, Version #1.30			
CC	CC	CURRENT APPLICATION DATA:			
CC	CC	APPLICATION NUMBER: US/08/951,733			
CC	CC	FILING DATE: 16-OCT-1997			
CC	CC	CLASSIFICATION: 435			
CC	CC	PRIOR APPLICATION DATA:			
CC	CC	APPLICATION NUMBER: US 08/873,039			
CC	CC	FILING DATE: 11-JUN-1997			
CC	CC	PRIOR APPLICATION DATA:			
CC	CC	APPLICATION NUMBER: US 08/751,189			
CC	CC	FILING DATE: 15-NOV-1996			
CC	CC	ATTORNEY/AGENT INFORMATION:			
CC	CC	NAME: Oleski, Nancy A.			
CC	CC	REGISTRATION NUMBER: 34,688			
CC	CC	REFERENCE/DOCKET NUMBER: A-433B			
CC	CC	TELECOMMUNICATION INFORMATION:			
CC	CC	TELEPHONE: (805) 447-6504			
CC	CC	TELEFAX: (805) 499-8011			
CC	CC	INFORMATION FOR SEQ ID NO: 20:			

CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1154 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: unknown  
CC TOPOLOGY: unknown  
CC MOLECULE TYPE: protein  
CC SEQUENCE 1154 AA; 129526 MW; 6842246 CN

Query Match	100.0%;	Score 8624;	DB 13;	Length 1154;
Best Local Similarity	100.0%;	Pred. No. 0.00e+00;		
Matches 1154; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

D	b	1	HASORCVLLJBTWELALPAPFAMPARARCAVBSLESSHREVLPLATFVRRJGPOGMRL	60
O	y	1	HASOQRCVLLJBTWELALPAPFAMPARPCRAVBSLSHREVLPLATFVRRJGPOGMRL	60
D	b	61	VORODPAAFRLVAVQCLVCMVMDARPPAPASFQSOVSCLELVARVLQORCEGANKVLA	120
O	y	61	VORODPAAFRLVAVQCLVCMVMDARPPAPASFQSOVSCLELVARVLQORCEGANKVLA	120
D	b	121	FGFALLDGAARGPPEAFTTSVRSYLPNTVTDALRGSGAMGILLRRVDDVLVHLARCAL	180
O	y	121	FGFALLDGAARGPPEAFTTSVRSYLPNTVTDALRGSGAMGILLRRVDDVLVHLARCAL	180
D	b	181	FVLVAPSCAYVOCGRPLYQLGALATQARPPRPAASPPRRRLGCEERAMNHSYBAGVPGJPA	240
O	y	181	FVLVAPSCAYVOCGRPLYQLGALAAQARPPRPAASPPRRRLGCEERAMNHSYBAGVPGJPA	240
D	b	241	PGARRGGSASRSJLPLPRPRRGAAPBERERTPVQOGSMARPGRRGSDGFCVSPARP	300
O	y	241	PGARRGGSASRSJLPLPRPRRGAAPBERERTPVQOGSMARPGRRGSDGFCVSPARP	300
D	b	301	AEEATSLLEGALSGTRHSHSPVGRONHAGPPSTSRPBRMDTPCPPVYAETKHFLYSSGDK	360
O	y	301	AEEATSLLEGALSGTRHSHSPVGRONHAGPPSTSRPBRMDTPCPPVYAETKHFLYSSGDK	360
D	b	361	EOLRPSFLSSLRSLRGARLYVETITLGSPPMPGTPRRJLRLPQRYOMRPLFLELIG	420
O	y	361	EOLRPSFLSSLRSLRGARLYVETITLGSPPMPGTPRRJLRLPQRYOMRPLFLELIG	420
D	b	421	NHAOCPPYVLKTCOPLDLAAATPAAGVCARCKPGGSAAPBEEDTDBRRJLYOLLROHSSP	480
O	y	421	NHAOCPPYVLKTCOPLDLAAATPAAGVCARCKPGGSAAPBEEDTDBRRJLYOLLROHSSP	480
D	b	481	MOVYGFVYACLRRLVPPGLMGSRRNERERFLNKKFTISLGKHAULTSLOELTWKMSYRDCA	540
O	y	481	MOVYGFVYACLRRLVPPGLMGSRRNERERFLNKKFTISLGKHAULTSLOELTWKMSYRDCA	540
D	b	541	WLRSPGVCYPAABHRLREBELLAKFLHMLMSYVVELNSFFVYETTTQKNRFLFYRK	600
O	y	541	WLRSPGVCYPAABHRLREBELLAKFLHMLMSYVVELNSFFVYETTTQKNRFLFYRK	600
D	b	601	SWMSLQSIGIRHOKRQVLELSEAEVROHREARPLTISRLEIFKPGGLRPIYMDY	660
O	y	601	SWMSLQSIGIRHOKRQVLELSEAEVROHREARPLTISRLEIFKPGGLRPIYMDY	660
D	b	661	VVGATFERERKRAELTISRVALTESVLYNEEARRRPGLLGASVGLDIDIRHAMPTFVLYR	720
O	y	661	VVGATFERERKRAELTISRVALTESVLYNEEARRRPGLLGASVGLDIDIRHAMPTFVLYR	720
D	b	721	AODPPELJYFKVVDVGTGYDTIPDDRLETVIASIIRKQNTYCVRRVAVOKAAGHVRKA	780
O	y	721	AODPPELJYFKVVDVGTGYDTIPDDRLETVIASIIRKQNTYCVRRVAVOKAAGHVRKA	780
D	b	781	FKSHVSTLTDQPYRKQVVAHQETSPLRDVAVIEGOSSLSNEASSGLFVFLRPMCHNAV	840
O	y	781	FKSHVSTLTDQPYRKQVVAHQETSPLRDVAVIEGOSSLSNEASSGLFVFLRPMCHNAV	840
D	b	841	RIRKGSYVQOCGIRPOGSLTSLTSLCAGDENKTLFGIRRDGILLRLYDDELVLVPHLT	900
O	y	841	RIRKGSYVQOCGIRPOGSLTSLTSLCAGDENKTLFGIRRDGILLRLYDDELVLVPHLT	900
D	b	901	HAFTFLRLVAGVEYEGCVNLRKTVNFPEDBALGGTAFVOMPAHGLFPMCGILLDTR	960

Qy	901	HAKEFTLTVRGVPEYGCYVNLAKTVNVPVEDELAGTAFOHABHLEFMCGLLIDTR	960
Db	961	TLEQDSYSSYATSTRASITFRNGFKAGRNMRKLFGLRLKCHSLFLDDQVNSLOTVC	1020
Qy	961	TLEQDSYSSYATSTRASITFRNGFKAGRNMRKLFGLRLKCHSLFLDDQVNSLOTVC	1020
Db	1021	TNIYKILLQAFRFHACVQLDPRHQOVMKNPTEFLRVISDTASLCYSTLKAKNAGMSGIGA	1080
Qy	1021	TNIYKILLQAFRFHACVQLDPRHQOVMKNPTEFLRVISDTASLCYSTLKAKNAGMSGIGA	1080
Db	1081	KGAAGPLPSEAVQWMLCHOAFILKLTFRHRTVYVPLGLSRTAQQLSRKLPSTTTLAEAA	1140
Qy	1081	KGAAGPLPSEAVQWMLCHOAFILKLTFRHRTVYVPLGLSRTAQQLSRKLPSTTTLAEAA	1140
Db	1141	ANPALPBSDFKTIID	1154
Qy	1141	ANPALPBSDFKTIID	1154

RESULT	2		
ID	US-08-974-613	STANDARD;	PRT: 1189 AA.
XX			
AC	xxxxxx		
XX			
DT			
DE	Sequence 613, Application US/08974549		
XX			
CC	Sequence 613, Application US/08974549		
CC	GENERAL INFORMATION:		
CC	APPLICANT: Cecch, Thomas R.		
CC	APPLICANT: Lingner, Joachim		
CC	APPLICANT: Nakamura, Toru		
CC	APPLICANT: Chapman, Karen B.		
CC	APPLICANT: Morin, Gregg B.		
CC	APPLICANT: Hartley, Calvin B.		
CC	APPLICANT: Andrews, William H.		
CC	TITLE OF INVENTION: Human Telomerase Catalytic Subunit		
CC	NUMBER OF SEQUENCES: 726		
CC	CORRESPONDENCE ADDRESSES:		
CC	ADDRESSEE: Townsend and Townsend and Crew LLP		
CC	STREET: Two Embarcadero Center, Eighth Floor		
CC	CITY: San Francisco		
CC	STATE: California		
CC	COUNTRY: USA		
CC	ZIP: 94111-3834		
CC	COMPUTER READABLE FORM:		
CC	MEDIUM TYPE: Floppy disk		
CC	COMPUTER: IBM PC compatible		
CC	OPERATING SYSTEM: PC-DOS/MS-DOS		
CC	SOFTWARE: PatentIn Release #1.0, Version #1.30		
CC	CURRENT APPLICATION DATA:		
CC	APPLICATION NUMBER: US/08/974,549		
CC	FILING DATE: 19-NOV-1997		
CC	CLASSIFICATION: 536		
CC	PRIOR APPLICATION DATA:		
CC	APPLICATION NUMBER: US 08/724,643		
CC	FILING DATE: 01-OCT-1996		
CC	PRIOR APPLICATION DATA:		
CC	APPLICATION NUMBER: US 08/844,419		
CC	FILING DATE: 18-APR-1997		
CC	PRIOR APPLICATION DATA:		
CC	APPLICATION NUMBER: US 08/846,017		
CC	FILING DATE: 25-APR-1997		
CC	PRIOR APPLICATION DATA:		
CC	APPLICATION NUMBER: US 08/851,843		
CC	FILING DATE: 06-MAY-1997		
CC	PRIOR APPLICATION DATA:		
CC	APPLICATION NUMBER: US 08/854,050		
CC	FILING DATE: 09-MAY-1997		
CC	PRIOR APPLICATION DATA:		
CC	APPLICATION NUMBER: US 08/911,312		



```

CC      ZIP: 94111-3834
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Floppy disk
CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: Patentin Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: US/08/911,312
CC      FILING DATE: 14-AUG-1997
CC      CLASSIFICATION: 536
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: US 08/724,643
CC      FILING DATE: 01-OCT-1996
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: US 08/844,419
CC      FILING DATE: 18-APR-1997
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: US 08/846,017
CC      FILING DATE: 25-APR-1997
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: US 08/851,843
CC      FILING DATE: 06-MAY-1997
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: US 08/854,050
CC      FILING DATE: 09-MAY-1997
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: US 08/912,951
CC      FILING DATE: 14-AUG-1997
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: US 08/915,503
CC      FILING DATE: 14-AUG-1997
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: Elmhorn, Gregory P.
CC      REGISTRATION NUMBER: 36,440
CC      REFERENCE/DOCKET NUMBER: 015369-002500US
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (415) 576-0200
CC      TELEFAX: (415) 576-0300
CC      INFORMATION FOR SEQ ID NO: 34:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 1189 amino acids
CC      TYPE: amino acid
CC      STRANDEDNESS:
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: protein
CC      SEQUENCE 1189 AA; 133179 MW; 7256545 CN;
SQ
Query Match      99.8%; Score 8607; DB 13; Length 1189;
Best Local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 1152; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      302  EATSLGALSIGTRSHSHSVGROHAGPPSTSRPPRMDTPCPVYAETKHFLYSSGDKE 361
DB      397  QLRSEFLLSLRPSLTGARRLVERIFLGSRRPMGTRRLPLRQRMWRPLLELLGN 456
QY      362  QLRSEFLLSLRPSLTGARRLVERIFLGSRRPMGTRRLPLRQRMWRPLLELLGN 421
DB      457  HAQCPYGLTKHCPPLRAAYTPAAGVCAREKPOGGSVAPEEEDDPRLVQLLRQHSPP 516
QY      422  HAQCPYGLTKHCPPLRAAYTPAAGVCAREKPOGGSVAPEEEDDPRLVQLLRQHSPP 481
DB      517  QVYGFVRACLRRLVPPGLWSRHNERFLRNTKFFISLGHAKLSLOELTKMSVRDCAW 576
QY      482  QVYGFVRACLRRLVPPGLWSRHNERFLRNTKFFISLGHAKLSLOELTKMSVRDCAW 541
DB      577  LRRSPGVGCVPAAHRLREELIAFLWMSVYVELLSRFYTTETTPQKNRFFYRKS 636
QY      542  LRRSPGVGCVPAAHRLREELIAFLWMSVYVELLSRFYTTETTPQKNRFFYRKS 601
DB      637  VMSKLSIGIRQHLKRVQLRELSAEVROHREARPALITSRLRFIPPDGLRPIVNDYV 696
QY      602  VMSKLSIGIRQHLKRVQLRELSAEVROHREARPALITSRLRFIPPDGLRPIVNDYV 661
DB      697  VGARTFRERKRAERLTSRVKALFSLVNYERARRPGLGASVGLDIIHRMRFFVLRYRA 756
QY      662  VGARTFRERKRAERLTSRVKALFSLVNYERARRPGLGASVGLDIIHRMRFFVLRYRA 721
DB      757  QDPPELYFVAVDYTGAYDITPDRLTEVIASTIKPONTICVRRYAVYQRAAHGVKAF 816
QY      722  QDPPELYFVAVDYTGAYDITPDRLTEVIASTIKPONTICVRRYAVYQRAAHGVKAF 781
DB      817  KSHVSTLTDLPYKQVFAHLOETSPLRDAVIEQSSSLNEASGLPDVFLREWCNHAVR 876
QY      782  KSHVSTLTDLPYKQVFAHLOETSPLRDAVIEQSSSLNEASGLPDVFLREWCNHAVR 841
DB      877  IRGKSYVOCQGIPOGSLSTLSCYGDMEENKLFAGIRBDGLLRVDDFLVTPHLTH 936
QY      842  IRGKSYVOCQGIPOGSLSTLSCYGDMEENKLFAGIRBDGLLRVDDFLVTPHLTH 901
DB      937  AKTLRLTLVRGVPEYGCYVNLKRTVNFPEVDEALGTAFOQAHGLFPGCGLLDTRT 996
QY      902  AKTLRLTLVRGVPEYGCYVNLKRTVNFPEVDEALGTAFOQAHGLFPGCGLLDTRT 961
DB      997  LEVOSDYSYARTSIRASLTFNFGFKAGRNRRKLFVLRKCHSLDLDOVNSLOQVCT 1056
QY      962  LEVOSDYSYARTSIRASLTFNFGFKAGRNRRKLFVLRKCHSLDLDOVNSLOQVCT 1021
DB      1057  NIKYKILLQAYRFHACVYLQEPFHQVWKNPTFFLRVISTDASLICYSLKAKNAGMSIGAK 1116
QY      1022  NIKYKILLQAYRFHACVYLQEPFHQVWKNPTFFLRVISTDASLICYSLKAKNAGMSIGAK 1081
DB      1117  GAAGPLPSEAVQMLCHQAFLLKLRHRYTVYPLLSGSLRTAQTQSLRKLPGTTLLALPAA 1176
QY      1082  GAAGPLPSEAVQMLCHQAFLLKLRHRYTVYPLLSGSLRTAQTQSLRKLPGTTLLALPAA 1141
DB      1177  NPALPSPDEKTIID 1189
QY      1142  NPALPSPDEKTIID 1154

```

```

RESULT      4
ID          US-08-912-951-325      STANDARD;      PRT;      1189 AA.
XX          xxxxxx
AC          xxxxxx
XX          xxxxxx
DT          xxxxxx
XX          xxxxxx
DE          Sequence 325, Application US/08912951
CC          GENERAL INFORMATION:
CC          APPLICANT: Cech, Thomas R.
CC          APPLICANT: Lingner, Joachim
CC          APPLICANT: Nakamura, Toru

```



XX  
AC xxxxxx  
XX  
DE  
XX  
XX  
Sequence 612, Application US/08974549  
CC  
XX Sequence 612, Application US/08974549  
CC  
GENERAL INFORMATION:  
CC  
APPLICANT: Cech, Thomas R.  
CC  
APPLICANT: Lingner, Joachim  
CC  
APPLICANT: Nakamura, Toru  
CC  
APPLICANT: Chapman, Karen B.  
CC  
APPLICANT: Morlin, Gregg B.  
CC  
APPLICANT: Hartley, Calvin B.  
CC  
APPLICANT: Andrews, William H.  
CC  
TITLE OF INVENTION: Human Telomerase Catalytic Subunit  
CC  
NUMBER OF SEQUENCES: 726  
CC  
CORRESPONDENCE ADDRESS:  
CC  
ADDRESSEE: Townsend and Townsend and Crew LLP  
CC  
STREET: Two Embarcadero Center, Eighth Floor  
CC  
CITY: San Francisco  
CC  
STATE: California  
CC  
COUNTRY: USA  
CC  
ZIP: 94111-3834  
CC  
COMPUTER READABLE FORM:  
CC  
MEDIUM TYPE: Floppy disk  
CC  
COMPUTER: IBM PC compatible  
CC  
OPERATING SYSTEM: PC-DOS/MS-DOS  
CC  
SOFTWARE: Patent Release #1.0, Version #1.30  
CC  
CURRENT APPLICATION DATA:  
CC  
APPLICATION NUMBER: US/08/974,549  
CC  
FILING DATE: 19-NOV-1997  
CC  
CLASSIFICATION: 536  
CC  
PRIOR APPLICATION DATA:  
CC  
APPLICATION NUMBER: US 08/724,643  
CC  
FILING DATE: 01-OCT-1996  
CC  
PRIOR APPLICATION DATA:  
CC  
APPLICATION NUMBER: US 08/844,419  
CC  
FILING DATE: 18-APR-1997  
CC  
PRIOR APPLICATION DATA:  
CC  
APPLICATION NUMBER: US 08/846,017  
CC  
FILING DATE: 25-APR-1997  
CC  
PRIOR APPLICATION DATA:  
CC  
APPLICATION NUMBER: US 08/851,843  
CC  
FILING DATE: 06-MAY-1997  
CC  
PRIOR APPLICATION DATA:  
CC  
APPLICATION NUMBER: US 08/854,050  
CC  
FILING DATE: 09-MAY-1997  
CC  
PRIOR APPLICATION DATA:  
CC  
APPLICATION NUMBER: US 08/911,312  
CC  
FILING DATE: 14-AUG-1997  
CC  
PRIOR APPLICATION DATA:  
CC  
APPLICATION NUMBER: US 08/912,951  
CC  
FILING DATE: 14-AUG-1997  
CC  
PRIOR APPLICATION DATA:  
CC  
APPLICATION NUMBER: US 08/915,503  
CC  
FILING DATE: 14-AUG-1997  
CC  
PRIOR APPLICATION DATA:  
CC  
APPLICATION NUMBER: WO PCT/US97/17618  
CC  
FILING DATE: 01-OCT-1997  
CC  
PRIOR APPLICATION DATA:  
CC  
APPLICATION NUMBER: WO PCT/US97/17865  
CC  
FILING DATE: 01-OCT-1997  
CC  
ATTORNEY/AGENT INFORMATION:  
CC  
NAME: Apple, Randolph Ted  
CC  
REGISTRATION NUMBER: 36,429  
CC  
REFERENCE/DOCKET NUMBER: 015389-002610US  
CC  
TELECOMMUNICATION INFORMATION:  
CC  
TELEPHONE: (415) 576-0200  
CC  
TELEFAX: (415) 576-0300  
CC  
INFORMATION FOR SEQ ID NO: 612:  
CC  
SEQUENCE CHARACTERISTICS:

CC LENGTH: 1200 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS:  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC FEATURE:  
CC NAME/KEY: Protein  
CC LOCATION: 1..1200  
CC OTHER INFORMATION: /note="fusion protein composed of His6  
CC OTHER INFORMATION: and Anti-Xpress tags, enterokinase  
CC OTHER INFORMATION: cleavage site and full length hprt  
CC OTHER INFORMATION: protein"  
SQ SEQUENCE 1200 AA: 134322 MW: 7387257 CN:  
  
Query Match 99.88; Score 8607; DB 18; Length 1200;  
Best Local Similarity 99.98; Pred. No. 0.00e+00;  
Matches 1152; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Db 48 ASTGCVLTWEALAPATPAMPAPRCRAVRSLSRSHYREVLPATFVRLAPGQWRV 107  
|||  
Qy 2 ASGRCVLTWEALAPATPAMPAPRCRAVRSLSRSHYREVLPATFVRLAPGQWRV 61  
  
Db 108 QRGDPAFRLVAQCLVCPWMDAPPAPPSFROVSCLELVARVYLQRLCEGAKNYLAF 167  
|||  
Qy 62 QRGDPAFRLVAQCLVCPWMDAPPAPPSFROVSCLELVARVYLQRLCEGAKNYLAF 121  
  
Db 168 GFALIDGARGPPPAFTTSTVSTILPNTVDALGSGAMGILLRVDVYVHLARCALF 227  
|||  
Qy 122 GFALIDGARGPPPAFTTSTVSTILPNTVDALGSGAMGILLRVDVYVHLARCALF 181  
  
Db 228 VLVAPSCAYOVCGPPLVQLAATQARPPHASPGRRLGGERAMNSVRSAGVPLGLPAP 287  
|||  
Qy 182 VLVAPSCAYOVCGPPLVQLAATQARPPHASPGRRLGGERAMNSVRSAGVPLGLPAP 241  
  
Db 288 GARRRGASASRLPLPRPRRGAAPPEERTPVQGSWAHPDRTGSDRGFCVSPARPA 347  
|||  
Qy 242 GARRRGASASRLPLPRPRRGAAPPEERTPVQGSWAHPDRTGSDRGFCVSPARPA 301  
  
Db 348 EEATSLGALSGTRHSHSPSVGRHHAGPSTSRPPRWDTPCCPPVYAEYHFLYSSDKE 407  
|||  
Qy 302 EEATSLGALSGTRHSHSPSVGRHHAGPSTSRPPRWDTPCCPPVYAEYHFLYSSDKE 361  
  
Db 408 QLRPSFLSSLRPSLTGARLVETIFLGSRRMPGFRRLPRLPQRWQMRPFLLELGN 467  
|||  
Qy 362 QLRPSFLSSLRPSLTGARLVETIFLGSRRMPGFRRLPRLPQRWQMRPFLLELGN 421  
  
Db 468 HAQCPYVLLKTHCPRLAAVTPAAGVCAREKPGQSVAAPEEDTDPRLVOLLRQSSPW 527  
|||  
Qy 422 HAQCPYVLLKTHCPRLAAVTPAAGVCAREKPGQSVAAPEEDTDPRLVOLLRQSSPW 481  
  
Db 528 QYGFVYACLRRLVPPGLMGSRHNERFLNRTKFTISLGHAKLSIQELTWKGSYRDCAF 587  
|||  
Qy 482 QYGFVYACLRRLVPPGLMGSRHNERFLNRTKFTISLGHAKLSIQELTWKGSYRDCAF 541  
  
Db 588 LRSRPGVGVPAAEHRLREELIKFLHMLSVYVELLRSEFFVYETTFQKNLFFYRKS 647  
|||  
Qy 542 LRSRPGVGVPAAEHRLREELIKFLHMLSVYVELLRSEFFVYETTFQKNLFFYRKS 601  
  
Db 648 VWSKLSIGIRHILKRVQLRELSEAEVQRHREARPALTSRLFTIPKDLRPIVNNDDV 707  
|||  
Qy 602 VWSKLSIGIRHILKRVQLRELSEAEVQRHREARPALTSRLFTIPKDLRPIVNNDDV 661  
  
Db 708 VGATFRERKRAERLISRVKALFSLYNTYEPARPPGLIGASVGLGDIHRAWRFFVLRVA 767  
|||  
Qy 662 VGATFRERKRAERLISRVKALFSLYNTYEPARPPGLIGASVGLGDIHRAWRFFVLRVA 721  
  
Db 768 QDPEPELYFVKVDVTGAYVDITIPDRLTEVASTIKPONTCVRRYAVYQAAAGHYRKAF 827  
|||  
Qy 722 QDPEPELYFVKVDVTGAYVDITIPDRLTEVASTIKPONTCVRRYAVYQAAAGHYRKAF 781  
  
Db 828 KSHVSTLTDLPYRQFVAMQETSPLRDVAVIEQSSSLNEASSGLEDVFLRPMCHRAVR 887  
|||  
Qy 782 KSHVSTLTDLPYRQFVAMQETSPLRDVAVIEQSSSLNEASSGLEDVFLRPMCHRAVR 841

DB 888 IRGKSYVOCGIPGSIITLILGSLCYGDMENKLFAGIRRDGLLRVLVDPLVPHLTH 947  
CC  
CC  
QY 842 IRGKSYVOCGIPGSIITLILGSLCYGDMENKLFAGIRRDGLLRVLVDPLVPHLTH 901  
CC  
DB 948 AKTFLRTLVRCVPEVGCYNLRKTYVNFVDEALGTAFAVOMPAHGLFPMCGLLDTRT 1007  
CC  
QY 902 AKTFLRTLVRCVPEVGCYNLRKTYVNFVDEALGTAFAVOMPAHGLFPMCGLLDTRT 961  
CC  
DB 1008 LEVOSDYSSVARTSIRASLTFNRGFKAGRNMRKLFVLRKCHSLFLDLQVNSLQTVCT 1067  
CC  
QY 962 LEVOSDYSSVARTSIRASLTFNRGFKAGRNMRKLFVLRKCHSLFLDLQVNSLQTVCT 1021  
CC  
DB 1068 NIKYLLILQARFACVQLVLPFHQVWKNPTEFLRVISDTLSLCTSIKANNAGSLGAK 1127  
CC  
QY 1022 NIKYLLILQARFACVQLVLPFHQVWKNPTEFLRVISDTLSLCTSIKANNAGSLGAK 1081  
CC  
DB 1128 GAAGPLPSEAVOMLCHQAFLLKTRHRTYVPLGLSLRTAOTQLSRKLPGLTTLALEAA 1187  
CC  
QY 1082 GAAGPLPSEAVOMLCHQAFLLKTRHRTYVPLGLSLRTAOTQLSRKLPGLTTLALEAA 1141  
CC  
DB 1188 NPALPSDEKTIID 1200  
CC  
QY 1142 NPALPSDEKTIID 1154  
CC  
RESULT 6 US-08-912-951-324 STANDARD; PRT; 1200 AA.  
ID US-08-912-951-324 STANDARD; PRT; 1200 AA.  
AC xxxxxx  
XX  
DE Sequence 324, Application US/08912951  
XX  
CC GENERAL INFORMATION:  
CC  
CC APPLICANT: Cech, Thomas R.  
CC APPLICANT: Lingner, Joachim  
CC APPLICANT: Nakamura, Toru  
CC APPLICANT: Chapman, Karen B.  
CC APPLICANT: Morlin, Gregg B.  
CC APPLICANT: Harley, Calvin  
CC APPLICANT: Andrews, William H.  
CC TITLE OF INVENTION: HUMAN TELOMERASE CATALYTIC SUBUNIT: DIAGNOSTIC AND  
CC TITLE OF INVENTION: THERAPEUTIC METHODS  
CC NUMBER OF SEQUENCES: 335  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Townsend and Townsend and Crew LLP  
CC STREET: Two Embarcadero Center, 8th Floor  
CC CITY: San Francisco  
CC STATE: California  
CC COUNTRY: United States of America  
CC ZIP: 94111  
CC  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: IBM PC compatible  
CC SOFTWARE: Patent Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/912,951  
CC FILING DATE: 14-AUG-1997  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/854,050  
CC FILING DATE: 09-MAY-1997  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/851,843  
CC FILING DATE: 06-MAY-1997  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/846,017

CC FILING DATE: 25-APR-1997  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/844,419  
CC FILING DATE: 18-APR-1997  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/724,643  
CC FILING DATE: 01-OCT-1996  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Apple, Randolph T.  
CC REGISTRATION NUMBER: 36,429  
CC REFERENCE/DOCKET NUMBER: 015389-002600US  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 576-0200  
CC TELEFAX: (415) 576-0300  
CC INFORMATION FOR SEQ ID NO: 324:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1200 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS:  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 1200 AA; 134322 MW; 7387257 CN;  
SQ  
Query Match 99.8%; Score 8607; DB 13; Length 1200;  
Best Local Similarity 99.9%; Pred. No. 0.00e+00;  
Matches 1152; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
DB 48 ASTORCVLRTWELALPATPMRPARPCRAVRSLLRSHYREVLPLATFVRRLGPGGRVLY 107  
CC  
QY 2 ASGRCVLTWELALPATPMRPARPCRAVRSLLRSHYREVLPLATFVRRLGPGGRVLY 61  
CC  
DB 108 QRGDPAFRAVLAOCVLPWDARPPAASFRQVSLKELVAVYLQRLCEGAKNVLA 167  
CC  
QY 62 QRGDPAFRAVLAOCVLPWDARPPAASFRQVSLKELVAVYLQRLCEGAKNVLA 121  
CC  
DB 168 GFALLDARGGPPPAFTTSVRSYLPNTVTDALRGSGAMGLLRVGDVYLHLARCALF 227  
CC  
QY 122 GFALLDARGGPPPAFTTSVRSYLPNTVTDALRGSGAMGLLRVGDVYLHLARCALF 181  
CC  
DB 228 VLVAPSCAYVCGPPVQLGAAQTARPPPHASGPRRLGCEPRAVNHVSREAGVPLGAP 287  
CC  
QY 182 VLVAPSCAYVCGPPVQLGAAQTARPPPHASGPRRLGCEPRAVNHVSREAGVPLGAP 241  
CC  
DB 288 GARRRGASASRLPLPRKPRGAPEPERPVGGGSAHPGRTGSPDRGCVVSPARA 347  
CC  
QY 242 GARRRGASASRLPLPRKPRGAPEPERPVGGGSAHPGRTGSPDRGCVVSPARA 301  
CC  
DB 348 EEAATSLGALSGTRHSHPSVGRQHHAGPSTSRPPRMDTPCPVVAETKHFILYSSGDK 407  
CC  
QY 302 EEAATSLGALSGTRHSHPSVGRQHHAGPSTSRPPRMDTPCPVVAETKHFILYSSGDK 361  
CC  
DB 408 QLRPSFLISSLRBSLTGARLVETIFLGSRPWMPCTRRRLRLQRTWQMPRLLELLGN 467  
CC  
QY 362 QLRPSFLISSLRBSLTGARLVETIFLGSRPWMPCTRRRLRLQRTWQMPRLLELLGN 421  
CC  
DB 468 HAQCPYGLTKHCPRLRAVTPPAAGVCAREKPGQSVAAPEEDTDPRRLVOLLRQHSPPW 527  
CC  
QY 422 HAQCPYGLTKHCPRLRAVTPPAAGVCAREKPGQSVAAPEEDTDPRRLVOLLRQHSPPW 481  
CC  
DB 528 QYGFVTRACLRVLPGLWGRHNERRLRNTKFFISLGRKAKLSLOELTWKMSVRCDAW 587  
CC  
QY 482 QYGFVTRACLRVLPGLWGRHNERRLRNTKFFISLGRKAKLSLOELTWKMSVRCDAW 541  
CC  
DB 588 LRSPGVGCYVAAHRLREETLAKFLWIMSVYVVELLRSEFYTEETFOKNRLEFYRKS 647  
CC  
QY 542 LRSPGVGCYVAAHRLREETLAKFLWIMSVYVVELLRSEFYTEETFOKNRLEFYRKS 601  
CC  
DB 648 VMSKIOSIGIRQHLKRVOLRELSEAEYQRHREARPALITSLRLEIPRPDGLRPVNNDYV 707  
CC  
QY 602 VMSKIOSIGIRQHLKRVOLRELSEAEYQRHREARPALITSLRLEIPRPDGLRPVNNDYV 661  
CC

[illegible]

Cc		CLASSIFICATION:	536
Cc		PRIOR APPLICATION DATA:	
Cc		APPLICATION NUMBER:	US 08/724,643
Cc		FILING DATE:	01-OCT-1996
Cc		PRIOR APPLICATION DATA:	
Cc		APPLICATION NUMBER:	US 08/844,419
Cc		FILING DATE:	18-APR-1997
Cc		PRIOR APPLICATION DATA:	
Cc		APPLICATION NUMBER:	US 08/846,017
Cc		FILING DATE:	25-APR-1997
Cc		PRIOR APPLICATION DATA:	
Cc		APPLICATION NUMBER:	US 08/851,843
Cc		FILING DATE:	06-MAY-1997
Cc		PRIOR APPLICATION DATA:	
Cc		APPLICATION NUMBER:	US 08/854,050
Cc		FILING DATE:	09-MAY-1997
Cc		PRIOR APPLICATION DATA:	
Cc		APPLICATION NUMBER:	US 08/912,951
Cc		FILING DATE:	14-AUG-1997
Cc		PRIOR APPLICATION DATA:	
Cc		APPLICATION NUMBER:	US 08/915,503
Cc		FILING DATE:	14-AUG-1997
Cc		ATTORNEY/AGENT INFORMATION:	
Cc		NAME:	Einhorn, Gregory P.
Cc		REGISTRATION NUMBER:	38,440
Cc		TELEPHONE:	(415) 576-0200
Cc		TELEFAX:	(415) 576-0300
Cc		INFORMATION FOR SEQ ID NO:	33:
Cc		SEQUENCE CHARACTERISTICS:	
Cc		LENGTH:	1200 amino acids
Cc		TYPE:	amino acid
Cc		STRANDEDNESS:	
Cc		TOPOLOGY:	linear
Cc		MOLECULE TYPE:	protein
Cc		SEQUENCE	1200 AA: 134322 MM; 7387257 CN;
SQ			
Dd		Query Match	99.8%; Score 8607; DB 13; Length 1200;
		Best Local Similarity	99.9%; Pred.No. 0.00e+00;
		Matches 1152; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
Dd		48	ASTGRCVILRTWELAAATPMPAPAPPCRAVRSLIRSHYREVLPLAFVYRLDPQGRLV 107
Qy		2	ASGGRCVILRTWEALAAATPMPAPAPPCRAVRSLIRSHYREVLPLAFVYRLDPQGRLV 61
Dd		108	QRGPAAFRAIVACLVCPWDARPPPAASFROVSCLELVARVLOLRERGAKNYLAF 167
Qy		62	QRGPAAFRAIVACLVCPWDARPPPAASFROVSCLELVARVLOLRERGAKNYLAF 121
Dd		168	GFAILDGARGGPPEAFTTSVRSYLPNTVTDALRGSGAMGLLRVGDVLYVHLLARCALT 227
Qy		122	GFAILDGARGGPPEAFTTSVRSYLPNTVTDALRGSGAMGLLRVGDVLYVHLLARCALT 181
Dd		228	VLVAPSCAVQVCGRPLYQLGAATGARRRPPPAAGRRRLGGERAMNHSVVRAGVPLGIAP 287
Qy		182	VLVAPSCAVQVCGRPLYQLGAATGARRRPPPAAGRRRLGGERAMNHSVVRAGVPLGIAP 241
Dd		288	GARRGGASASLSLPRKPRRGAAPEDEPTPVGGGSMAHGRTGSPDRGFCCVVSAPARA 347
Qy		242	GARRGGASASLSLPRKPRRGAAPEDEPTPVGGGSMAHGRTGSPDRGFCCVVSAPARA 301
Dd		348	EEATSLLEGALSGTSHSPSVGRQHHAQPSTSRPPMDPPCPRVVYAETKHFLYSQDKE 407
Qy		302	EEATSLLEGALSGTSHSPSVGRQHHAQPSTSRPPMDPPCPRVVYAETKHFLYSQDKE 361
Dd		408	QLRPSTLIISLRPSLTGARRLVETIFLGSRPMWGTPRRRLPRLPQRYQMRLPFLELGN 467
Qy		362	QLRPSTLIISLRPSLTGARRLVETIFLGSRPMWGTPRRRLPRLPQRYQMRLPFLELGN 421
Dd		468	HAOCPTYGVLLKTHOPLAAATVPAGVCARAREKPGQSVAAPEEDTDPRRLVOLLRHSSPW 527
Qy		422	HAOCPTYGVLLKTHOPLAAATVPAGVCARAREKPGQSVAAPEEDTDPRRLVOLLRHSSPW 481

DB 528 OYGVFVACLRRLVPGILMGSRHNRRLRNKFTISLGNKAKLSIQELTWMKSVDCAW 587  
OY 482 OYGVFVACLRRLVPGILMGSRHNRRLRNKFTISLGNKAKLSIQELTWMKSVDCAW 541  
DB 588 LRSRSGVCVPAEAEHRLREELIAKFLHMLMSYVVELLSFFYVETETFOKRLFFYRS 647  
OY 542 LRSRSGVCVPAEAEHRLREELIAKFLHMLMSYVVELLSFFYVETETFOKRLFFYRS 601  
DB 648 VWSKQSIGIRHQLKRVOLRELSEAEVROHREARPALTSRLRFIPKPDGLRPINMDYV 707  
OY 602 VWSKQSIGIRHQLKRVOLRELSEAEVROHREARPALTSRLRFIPKPDGLRPINMDYV 661  
DB 708 VGARFRERKREKRLSRKALFVLYNTERARRPGLGVSGLDIDHMAWTFYLRYVA 767  
OY 662 VGARFRERKREKRLSRKALFVLYNTERARRPGLGVSGLDIDHMAWTFYLRYVA 721  
DB 768 QDPPELYEVKVDYTGAYDTIPQDLTEVIAIIRKQNTCYCRRYAVYQKAHGHVRAKAF 827  
OY 722 QDPPELYEVKVDYTGAYDTIPQDLTEVIAIIRKQNTCYCRRYAVYQKAHGHVRAKAF 781  
DB 828 KSHVSTLFDLQPYMROFVAHLQETSPLRDAVYIEQSSSLNEASSGLFVFLRFMCHAVR 887  
OY 782 KSHVSTLFDLQPYMROFVAHLQETSPLRDAVYIEQSSSLNEASSGLFVFLRFMCHAVR 841  
DB 888 IKGKSYVOCQIGIPQSSILSTLCSLCYGMENKLFAGIRDDLLRLVDDFLVPHLTH 947  
OY 842 IKGKSYVOCQIGIPQSSILSTLCSLCYGMENKLFAGIRDDLLRLVDDFLVPHLTH 901  
DB 948 AKTFLRTLVRGVPKGVNLRKTYVNFVEDEALGTAFAVQMPAHGLFPWCGLLDTRT 1007  
OY 902 AKTFLRTLVRGVPKGVNLRKTYVNFVEDEALGTAFAVQMPAHGLFPWCGLLDTRT 961  
DB 1008 LEVQDYSYASIRASLTFNRGFRAGNRMRKLEGVLRKCHSLFLLDLOVNSIQTVCT 1067  
OY 962 LEVQDYSYASIRASLTFNRGFRAGNRMRKLEGVLRKCHSLFLLDLOVNSIQTVCT 1021  
DB 1068 NIYKILLQAYFFHACVLDLPFHQOYWKMPTEFLRISTASLCSILKAKNAGSLGAK 1127  
OY 1022 NIYKILLQAYFFHACVLDLPFHQOYWKMPTEFLRISTASLCSILKAKNAGSLGAK 1081  
DB 1128 GAGGPLPSEAVOMLCHQAFLLKLRHRYVYVPLGSLRTAQTQSLKRLGTTLLALEAA 1187  
OY 1082 GAGGPLPSEAVOMLCHQAFLLKLRHRYVYVPLGSLRTAQTQSLKRLGTTLLALEAA 1141  
DB 1188 NPALPSDFKTIID 1200  
OY 1142 NPALPSDFKTIID 1154

RESULT 8  
ID US-08-974-549-600 STANDARD: PRT: 1285 AA.  
AC xxxxxx  
XX  
DT  
XX  
XX  
DE  
XX  
CC Sequence 600, Application US/08974549  
CC GENERAL INFORMATION:  
CC APPLICANT: Cech, Thomas R.  
CC APPLICANT: Lingner, Joachim  
CC APPLICANT: Nakamura, Toru  
CC APPLICANT: Chapman, Karen B.  
CC APPLICANT: Morin, Gregg B.  
CC APPLICANT: Hatley, Calvin B.  
CC APPLICANT: Andrews, William H.  
CC TITLE OF INVENTION: Human Telomerase Catalytic Subunit  
CC NUMBER OF SEQUENCES: 726  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Townsend and Townsend and Crew LLP  
CC STREET: Two Embarcadero Center, Eighth Floor

CC CITY: San Francisco  
CC STATE: California  
CC COUNTRY: USA  
CC ZIP: 94111-3834  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/974,549  
CC FILING DATE: 19-NOV-1997  
CC CLASSIFICATION: 536  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/724,643  
CC FILING DATE: 01-OCT-1996  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/844,419  
CC FILING DATE: 18-APR-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/846,017  
CC FILING DATE: 25-APR-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/851,843  
CC FILING DATE: 06-MAY-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/854,050  
CC FILING DATE: 09-MAY-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/911,312  
CC FILING DATE: 14-AUG-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/912,951  
CC FILING DATE: 14-AUG-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/915,503  
CC FILING DATE: 14-AUG-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: WO PCT/US97/17618  
CC FILING DATE: 01-OCT-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: WO PCT/US97/17885  
CC FILING DATE: 01-OCT-1997  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Apple, Randolph Ted  
CC REGISTRATION NUMBER: 36,429  
CC REFERENCE/DOCKET NUMBER: 015389-002610US  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 576-0200  
CC TELEFAX: (415) 576-0300  
CC INFORMATION FOR SEQ ID NO: 600:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1285 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS:  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC FEATURE:  
CC NAME/KEY: Protein  
CC LOCATION: 1..1285  
CC OTHER INFORMATION: /note- "fusion protein composed of  
CC OTHER INFORMATION: enterokinase cleavable, His tagged  
CC OTHER INFORMATION: thiorodoxin moiety and full length hTERT"  
CC  
SQ SEQUENCE 1285 AA; 143529 MW; 8449280 CN;

Query Match 99.8%; Score 8607; DB 18; Length 1285;  
Best Local Similarity 99.9%; Pred. No. 0.00e+00;  
Matches 1152; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 133 ASTORCVLLRTWEALAPATPAMPAPRCRAVRSLSRSHYREVLPATFVRRLGPOGWRVLY 192  
OY 2 ASGQCVLLRTWEALAPATPAMPAPRCRAVRSLSRSHYREVLPATFVRRLGPOGWRVLY 61

Db 193 QRGDPAERAIWAOCVCPMDARPPAPSPFQVSCLEKELAVRLOLRCERGANVLAF 252  
Qy 62 QRGDPAERAIWAOCVCPMDARPPAPSPFQVSCLEKELAVRLOLRCERGANVLAF 121  
Db 253 GFALIDGARGGPEAFTTSVRSYLPNTVTDALRGSGAGLLLRVDDVYLHLLARCALF 312  
Qy 122 GFALIDGARGGPEAFTTSVRSYLPNTVTDALRGSGAGLLLRVDDVYLHLLARCALF 181  
Db 313 VLVAPSCAYOVCPEPLVOLGAATOARPPHAGPBRRLGCEBAMHVSREAGVPLGLPAP 372  
Qy 182 VLVAPSCAYOVCPEPLVOLGAATOARPPHAGPBRRLGCEBAMHVSREAGVPLGLPAP 241  
Db 373 GARRRGASASRLPLPKRRPARGAEPERTPVQGSMAHPGTRGSPRGCEVSPAPPA 432  
Qy 242 GARRRGASASRLPLPKRRPARGAEPERTPVQGSMAHPGTRGSPRGCEVSPAPPA 301  
Db 433 EEAATSLGALSGTRSHSPSVGRQHAGPPSTSRPPMDTFCPPYATKHEFLYSSGDK 492  
Qy 302 EEAATSLGALSGTRSHSPSVGRQHAGPPSTSRPPMDTFCPPYATKHEFLYSSGDK 361  
Db 493 QLRBSFLISSRLPLTGARLVETIFLGSPPMGPTRRLPRLQRYOMPLFELLGN 552  
Qy 362 QLRBSFLISSRLPLTGARLVETIFLGSPPMGPTRRLPRLQRYOMPLFELLGN 421  
Db 553 HAQCPYGLTKHCPRLAAYTPAAGVCAREKPOGSVAPEEEDTPRRLVOLLRQHSSPM 612  
Qy 422 HAQCPYGLTKHCPRLAAYTPAAGVCAREKPOGSVAPEEEDTPRRLVOLLRQHSSPM 481  
Db 613 QYGVFVACLRRLVPPGLMGRHNERRLRNTKFTISLGNKAKLSLOELJWMSYRDCAM 672  
Qy 482 QYGVFVACLRRLVPPGLMGRHNERRLRNTKFTISLGNKAKLSLOELJWMSYRDCAM 541  
Db 673 LRSRGVGCVAAPAEERLAEELLAKFLHMLSVYVELLRSEFYVETEFQONRLEFYKRS 732  
Qy 542 LRSRGVGCVAAPAEERLAEELLAKFLHMLSVYVELLRSEFYVETEFQONRLEFYKRS 601  
Db 733 VMSKLOSIGIRHOLKRVOLRELSEAEVQRHREARPALITSRLRFIPKPDGLRPIYNDYV 792  
Qy 602 VMSKLOSIGIRHOLKRVOLRELSEAEVQRHREARPALITSRLRFIPKPDGLRPIYNDYV 661  
Db 793 VGARTFRERKAERLTSVKALFSYLVNTERARREGGLGASVGLDDIHRAMTFYLVRA 852  
Qy 662 VGARTFRERKAERLTSVKALFSYLVNTERARREGGLGASVGLDDIHRAMTFYLVRA 721  
Db 853 QDPPELYFVVDVDTGAVDTLPQDLTEVIAIITKQNTYCVRYAVYOKAHGVRRAF 912  
Qy 722 QDPPELYFVVDVDTGAVDTLPQDLTEVIAIITKQNTYCVRYAVYOKAHGVRRAF 781  
Db 913 KSHVSTLTDLPYMKROFVAHQETSPLRDAVYIEOSSSLNBAASSGLFDFVFLRFMCHNAVR 972  
Qy 782 KSHVSTLTDLPYMKROFVAHQETSPLRDAVYIEOSSSLNBAASSGLFDFVFLRFMCHNAVR 841  
Db 973 IRGKSYVOCGIPGOSIITSLCISLCYGDMEKNKLFAGIRRDGLLRVLDVDFLLVPHLTH 1032  
Qy 842 IRGKSYVOCGIPGOSIITSLCISLCYGDMEKNKLFAGIRRDGLLRVLDVDFLLVPHLTH 901  
Db 1033 AKTELRTLVKGVPEKGVVNNRKTYYNRPVDEALGCAFAVOMPAHGLFPWCGLLDRT 1092  
Qy 902 AKTELRTLVKGVPEKGVVNNRKTYYNRPVDEALGCAFAVOMPAHGLFPWCGLLDRT 961  
Db 1093 LEVOSDYSSYARTSIRASLTFRNGFKAGRNARRKLFGLVRLKCHSLFLDLVNSIQTYCT 1152  
Qy 962 LEVOSDYSSYARTSIRASLTFRNGFKAGRNARRKLFGLVRLKCHSLFLDLVNSIQTYCT 1021  
Db 1153 NIYKILLQAVRFHACVQLQEPFHQVWKNPTEFLRVISDTASLCYSILKANAGKSLGAK 1212  
Qy 1022 NIYKILLQAVRFHACVQLQEPFHQVWKNPTEFLRVISDTASLCYSILKANAGKSLGAK 1081  
Db 1213 GAAPPLSEAVOMLCHQAFLLKTRHRTYVPLGSLTAOTQSRKLPGLTTLRALEAA 1272  
Qy 1082 GAAPPLSEAVOMLCHQAFLLKTRHRTYVPLGSLTAOTQSRKLPGLTTLRALEAA 1141  
Db 1273 NPALPDSDFKTLID 1285

Qy 1142 NPALPDSDFKTLID 1154  
RESULT 9  
ID US-08-911-312-32 STANDARD: PRT: 1285 AA.  
AC xxxxxx  
XX  
XX  
XX  
DE Sequence 32, Application US/08911312  
XX  
CC Sequence 32, Application US/08911312  
CC GENERAL INFORMATION:  
CC APPLICANT: Cech, Thomas R.  
CC APPLICANT: Lingner, Joachim  
CC APPLICANT: Nakamura, Toru  
CC APPLICANT: Chapman, Karen B.  
CC APPLICANT: Morin, Gregg B.  
CC APPLICANT: Harley, Calvin B.  
CC APPLICANT: Andrews, William  
CC TITLE OF INVENTION: Telomerase Reverse Transcriptase  
CC NUMBER OF SEQUENCES: 170  
CC CORRESPONDENCE ADDRESSES:  
CC ADDRESSEE: Townsend and Townsend and Crew LLP  
CC STREET: Two Embarcadero Center, Eighth Floor  
CC CITY: San Francisco  
CC STATE: California  
CC COUNTRY: USA  
CC ZIP: 94111-3834  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/911,312  
CC FILING DATE: 14-AUG-1997  
CC CLASSIFICATION: 536  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/724,643  
CC FILING DATE: 01-OCT-1996  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/844,419  
CC FILING DATE: 18-APR-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/846,017  
CC FILING DATE: 25-APR-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/851,843  
CC FILING DATE: 06-MAY-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/854,050  
CC FILING DATE: 09-MAY-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/912,951  
CC FILING DATE: 14-AUG-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/915,503  
CC FILING DATE: 14-AUG-1997  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Einhorn, Gregory P.  
CC REGISTRATION NUMBER: 38,440  
CC REFERENCE/DOCKET NUMBER: 015389-002500US  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 576-0200  
CC TELEFAX: (415) 576-0300  
CC INFORMATION FOR SEQ ID NO: 32:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1285 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS:

CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 SO SEQUENCE 1285 AA: 1435280 MW: 8449280 CN:  
 Query Match 99.8%; Score 8607; DB 13; Length 1285;  
 Best Local Similarity 99.9%; Pred. No. 0.00e+00;  
 Matches 1152; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 133 ASTORCVLLRTWEALAPAMPAPRCRAVNSILRSYHREVLPLATVRRILGPGCMGLV 192  
 QY 2 ASGQCVLLRTWEALAPAMPAPRCRAVNSILRSYHREVLPLATVRRILGPGCMGLV 61  
 Db 193 QRGDPAFRAVLAVOCLVCPMDAPPPAPSPFQVSCLEKELVARIQRLCEGANVIAF 252  
 QY 62 QRGDPAFRAVLAVOCLVCPMDAPPPAPSPFQVSCLEKELVARIQRLCEGANVIAF 121  
 Db 253 GFALLDGAARGGPPFAFTTSVRSYLPNTVTDALRGSGAGMLLRVGDVYLHLLARCALF 312  
 QY 122 GFALLDGAARGGPPFAFTTSVRSYLPNTVTDALRGSGAGMLLRVGDVYLHLLARCALF 181  
 Db 313 VLVAPSCAYVCCGPPYLQGAATQARPPPHASGPPRRRLGCEPAMNHSYREAGVPLGLAP 372  
 QY 182 VLVAPSCAYVCCGPPYLQGAATQARPPPHASGPPRRRLGCEPAMNHSYREAGVPLGLAP 241  
 Db 373 GARRRGSGASRSLLPKRRPRGGAPEPERTPVGGGSMAPHGRTRGSDRGFCVSPARPA 432  
 QY 242 GARRRGSGASRSLLPKRRPRGGAPEPERTPVGGGSMAPHGRTRGSDRGFCVSPARPA 301  
 Db 433 BEATLEGALSTGRSHSPVGRQHHAGPPSTSRPPRMDTPCPPYAETKHPLSSGDKX 492  
 QY 302 BEATLEGALSTGRSHSPVGRQHHAGPPSTSRPPRMDTPCPPYAETKHPLSSGDKX 361  
 Db 493 QLRFSLLSLRPSLTGARLVETIFFLSRPMPCPTPRRLPLPRYQMRPLFELLGN 552  
 QY 362 QLRFSLLSLRPSLTGARLVETIFFLSRPMPCPTPRRLPLPRYQMRPLFELLGN 421  
 Db 553 HAQCYGVLKTHCPRLPAVPPAAGVCAKREKPGSVAAPEEDTDPRLVOLLRQHSWP 612  
 QY 422 HAQCYGVLKTHCPRLPAVPPAAGVCAKREKPGSVAAPEEDTDPRLVOLLRQHSWP 481  
 Db 613 QYGVFVARACLRRLVPPGMLSGRHNERRFLRNTKFTISLGHAKLSLOELTWMMSVDCAM 672  
 QY 482 QYGVFVARACLRRLVPPGMLSGRHNERRFLRNTKFTISLGHAKLSLOELTWMMSVDCAM 541  
 Db 673 LRSFSGVCVPAEHRLEBEILAKFLHMLSVYVELLSFEFVYVETTFQKNRLEFFYRKS 732  
 QY 542 LRSFSGVCVPAEHRLEBEILAKFLHMLSVYVELLSFEFVYVETTFQKNRLEFFYRKS 601  
 Db 733 VWSKIOSIGIRQHLKRVOLRELSAEVQHRRARPALLTSRLRFLPKPGGLRPVNMIV 792  
 QY 602 VWSKIOSIGIRQHLKRVOLRELSAEVQHRRARPALLTSRLRFLPKPGGLRPVNMIV 661  
 Db 793 VGARTFRREKRAERLTSRKALFSVLNTERARPPGLGASVGLDDIHRAMRTFLVRYA 852  
 QY 662 VGARTFRREKRAERLTSRKALFSVLNTERARPPGLGASVGLDDIHRAMRTFLVRYA 721  
 Db 853 QDPPELYFVKVYDTGAYDTIPQDRLTEVIAIIRPQNTCYRRYAVVQKAAGHYRKA 912  
 QY 722 QDPPELYFVKVYDTGAYDTIPQDRLTEVIAIIRPQNTCYRRYAVVQKAAGHYRKA 781  
 Db 913 KSHVSTLIDLOPYMROFVAHLOETSPLRVAVYIEOSSSLNENSSGLFDFVFLFKMCHHAR 972  
 QY 782 KSHVSTLIDLOPYMROFVAHLOETSPLRVAVYIEOSSSLNENSSGLFDFVFLFKMCHHAR 841  
 Db 973 IRGKSYVOCQGIPOGSIILSLCSLCYGMENKLFAGIRDDLLRLVDFLLVPHLTH 1032  
 QY 842 IRGKSYVOCQGIPOGSIILSLCSLCYGMENKLFAGIRDDLLRLVDFLLVPHLTH 901  
 Db 1033 AKTFLRTLVRGVPEYGCVVNLKRTVNFVEDEALGTAFAVOMPAHGLEPPMCGLLDTRT 1092  
 QY 902 AKTFLRTLVRGVPEYGCVVNLKRTVNFVEDEALGTAFAVOMPAHGLEPPMCGLLDTRT 961  
 Db 1093 LEVQSDSYSTARTSIRASLTFNRGFRAGNMRRKLFGLVRLKCHSLFLDLOVNSIQTYCT 1152

QY 962 LEVQSDSYSTARTSIRASLTFNRGFRAGNMRRKLFGLVRLKCHSLFLDLOVNSIQTYCT 1021  
 Db 1153 NITYKILLQAFRFAVYQLQPLFHQOVWKNPFELRVISDFTASLCYSTIKAKNAGMSIGAK 1212  
 QY 1022 NITYKILLQAFRFAVYQLQPLFHQOVWKNPFELRVISDFTASLCYSTIKAKNAGMSIGAK 1081  
 Db 1213 GAAGPLPSEAVQWICHQAFILKTRHRYTYVPLGSLRTAQTOLSRKLPQTTLALEMAA 1272  
 QY 1082 GAAGPLPSEAVQWICHQAFILKTRHRYTYVPLGSLRTAQTOLSRKLPQTTLALEMAA 1141  
 Db 1273 NPALPSDEFTILD 1285  
 QY 1142 NPALPSDEFTILD 1154

RESULT 10  
 ID US-08-912-951-314 STANDARD; PRT; 1285 AA.  
 AC xxxxxx  
 XX  
 XX  
 DE Sequence 314, Application US/08912951  
 CC  
 CC Sequence 314, Application US/08912951  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Cech, Thomas R.  
 CC APPLICANT: Lingner, Joachim  
 CC APPLICANT: Nakamura, Toru  
 CC APPLICANT: Chapman, Karen B.  
 CC APPLICANT: Morin, Gregg B.  
 CC APPLICANT: Harley, Calvin H.  
 CC APPLICANT: Andrews, William H.  
 CC TITLE OF INVENTION: HUMAN TELOMERASE CATALYTIC SUBUNIT: DIAGNOSTIC AND  
 CC TITLE OF INVENTION: THERAPEUTIC METHODS  
 CC NUMBER OF SEQUENCES: 335  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Townsend and Townsend and Crew LLP  
 CC STREET: Two Embarcadero Center, 8th Floor  
 CC CITY: San Francisco  
 CC STATE: California  
 CC COUNTRY: United States of America  
 CC ZIP: 94111  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/912,951  
 CC FILING DATE: 14-AUG-1997  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/854,050  
 CC FILING DATE: 09-MAY-1997  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/851,843  
 CC FILING DATE: 06-MAY-1997  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/846,017  
 CC FILING DATE: 25-APR-1997  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/844,419  
 CC FILING DATE: 18-APR-1997  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/724,643  
 CC FILING DATE: 01-OCT-1996  
 CC CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph T.  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 015389-002600US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 314:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1285 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE 1285 AA: 143529 MW: 8449280 CN:  
Query Match 99.8%; Score 8607; DB 13; Length 1285;  
Best Local Similarity 99.9%; Pred. No. 0.00e+00;  
Matches 1152; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 133 ASTORCVLLFTWEALAPATPAMPAPRCRAVRSILRSHTREYVPLATVRRLGPOGRLV 192  
QY 2 ASGRVCVLTWEALAPATPAMPAPRCRAVRSILRSHTREYVPLATVRRLGPOGRLV 61  
Db 193 QRGDPAFALVAOCIVCPMDARPPAPSFROYSCLEIYARVLCRGAKNVLA 252  
QY 62 QRGDPAFALVAOCIVCPMDARPPAPSFROYSCLEIYARVLCRGAKNVLA 121  
Db 253 GFALLDGAARGGPEAFTTSVRSYLPNTVTDALRSGANGLLRRYDDVLAHLCALF 312  
QY 122 GFALLDGAARGGPEAFTTSVRSYLPNTVTDALRSGANGLLRRYDDVLAHLCALF 181  
Db 313 VLVAPSCAYOVCSPYLOLGAATQARPPHAGRRRLGRCERAMHSHREGVPLGLAP 372  
QY 182 VLVAPSCAYOVCSPYLOLGAATQARPPHAGRRRLGRCERAMHSHREGVPLGLAP 241  
Db 373 GARRRGGSASRLPLPKPRRGAAPEPERTVYGOGSMHAPGTRGSPRGCVVSPARPA 432  
QY 242 GARRRGGSASRLPLPKPRRGAAPEPERTVYGOGSMHAPGTRGSPRGCVVSPARPA 301  
Db 433 EEAATSLGALSSTGRSHSHSVGRQHAGPPSTSRPPRPMDPCPPVYATKHFLLYSSGDK 492  
QY 302 EEAATSLGALSSTGRSHSHSVGRQHAGPPSTSRPPRPMDPCPPVYATKHFLLYSSGDK 361  
Db 493 QLRSPFLSLSRPLTGARLVETIFLGRPMWGTPTRLRLQRYQMPLLELLGN 552  
QY 362 QLRSPFLSLSRPLTGARLVETIFLGRPMWGTPTRLRLQRYQMPLLELLGN 421  
Db 553 HAQCPYGLTKHCPPLRAAVTPAAGVCAREKPOGSVAAPPEEDTDPRRLVOLLROHSSPW 612  
QY 422 HAQCPYGLTKHCPPLRAAVTPAAGVCAREKPOGSVAAPPEEDTDPRRLVOLLROHSSPW 481  
Db 613 QVYGFVRACTLRVLPGLMGRHNERFLRNTKFFISLGNHAKLSLOELTWKMSVRCDAW 672  
QY 482 QVYGFVRACTLRVLPGLMGRHNERFLRNTKFFISLGNHAKLSLOELTWKMSVRCDAW 541  
Db 673 LRRSPGVGCVAAHRLREELIAFLHMLSVYYVELLSRFYTTETTFQNNRLEFFTKS 732  
QY 542 LRRSPGVGCVAAHRLREELIAFLHMLSVYYVELLSRFYTTETTFQNNRLEFFTKS 601  
Db 733 VWSLQSGIGIQHQLKRVOLRELSEAEVQHREARPALITSLRFLPKPDGRLPIVNMNV 792  
QY 602 VWSLQSGIGIQHQLKRVOLRELSEAEVQHREARPALITSLRFLPKPDGRLPIVNMNV 661  
Db 793 VGARTFRERKRAERLTSVKALFSVLNTERARRBGLLGASVYLGDDIHRAARTFVLRYRA 852  
QY 662 VGARTFRERKRAERLTSVKALFSVLNTERARRBGLLGASVYLGDDIHRAARTFVLRYRA 721  
Db 853 QDPPELYFYKVVDTGAYDTIPQDRLETVLASIIPONTTVCYRRAYVQKAAHGHVRAAF 912  
QY 722 QDPPELYFYKVVDTGAYDTIPQDRLETVLASIIPONTTVCYRRAYVQKAAHGHVRAAF 781  
Db 913 KSHVSTLTDLPYKQFVAHLQETSPLDAVAVIQQSSSLNASSGLFDVFLRFMCHNAVR 972

|||||  
QY 782 KSHVSTLTDLPYKQFVAHLQETSPLDAVAVIQQSSSLNASSGLFDVFLRFMCHNAVR 841  
Db 973 IRGSIYVQCGIPIQGSILSTLLCSLCYGMENKLFACIRRDGLLRLVDDFLVTPHLTH 1032  
QY 842 IRGSIYVQCGIPIQGSILSTLLCSLCYGMENKLFACIRRDGLLRLVDDFLVTPHLTH 901  
Db 1033 AKTFLTLVGVPEYGVVNLKRVVNFPEDEALGSTAFVQMPAHGLFPMCGLLDTRT 1092  
QY 902 AKTFLTLVGVPEYGVVNLKRVVNFPEDEALGSTAFVQMPAHGLFPMCGLLDTRT 961  
Db 1093 LEVQSDYSYARTSIRASLTFFNRFKAGRMRRKLFGLRLKCHSLFLDLQVNSIQTVCT 1152  
QY 962 LEVQSDYSYARTSIRASLTFFNRFKAGRMRRKLFGLRLKCHSLFLDLQVNSIQTVCT 1021  
Db 1153 NITKILLQAYRFPACVQLPFIHQVWKNPFPLRVISDTASISLTKAKNAGMSIGAK 1212  
QY 1022 NITKILLQAYRFPACVQLPFIHQVWKNPFPLRVISDTASISLTKAKNAGMSIGAK 1081  
Db 1213 GAAGPLPSEAVQWICHOAFLTKLRHRTVYVPLGLSRTAOTLSRKLPGTTTLALEAA 1272  
QY 1082 GAAGPLPSEAVQWICHOAFLTKLRHRTVYVPLGLSRTAOTLSRKLPGTTTLALEAA 1141  
Db 1273 NPALPSDKTILD 1285  
QY 1142 NPALPSDKTILD 1154  
RESULT 11  
ID US-08-911-312-55 STANDARD; PRT: 1407 AA.  
AC xxxxxx  
XX  
XX  
Sequence 55, Application US/08911312  
DE  
XX  
CC Sequence 55, Application US/08911312  
CC GENERAL INFORMATION:  
CC APPLICANT: Cech, Thomas R.  
CC APPLICANT: Langner, Joachim  
CC APPLICANT: Nakamura, Toru  
CC APPLICANT: Chapman, Karen B.  
CC APPLICANT: Morin, Gregg B.  
CC APPLICANT: Hatley, Calvin B.  
CC APPLICANT: Andrews, William  
CC TITLE OF INVENTION: Telomerase Reverse Transcriptase  
CC NUMBER OF SEQUENCES: 170  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Townsend and Townsend and Crew LLP  
CC STREET: Two Embarcadero Center, Eighth Floor  
CC CITY: San Francisco  
CC STATE: California  
CC COUNTRY: USA  
CC ZIP: 94111-3834  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/911,312  
CC FILING DATE: 14-AUG-1997  
CC CLASSIFICATION: 536  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/724,643  
CC FILING DATE: 01-OCT-1996  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/844,419  
CC FILING DATE: 18-APR-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/846,017  
CC FILING DATE: 25-APR-1997

[illegible]

QY	602	VMSLQSGISGRQHLKRRQLEISEAEVYRQHRKRPMLLSRLRFLPKPDGLRPVYNDYV	661
Db	915	VGATFRERKRAEKLTSRVKALFESVLYTERARRPGLLGASVGLGDIHRAMPTFVLRYRA	974
OY	662	VGATFRERKRAEKLTSRVKALFESVLYTERARRPGLLGASVGLGDIHRAMPTFVLRYRA	721
Db	975	QDPPELFLFKVADYTGAYDITIPDDRLETVASITIKQNTCYVRRYAVVOKAAHGVRAAF	1033
OY	722	QDPPELFLFKVADYTGAYDITIPDDRLETVASITIKPONTYCVRRYAVVOKAAHGVRAAF	781
Db	1035	KSHTSLTDLOPYRVROVAHLOETSPLRDAVYVEOSSITNEASSGLFDEVLEFMCHEHVR	1094
OY	782	KSHTSLTDLOPYRVROVAHLOETSPLRDAVYVEOSSITNEASSGLFDEVLEFMCHEHVR	841
Db	1095	IRGSYVQOCGIPGSGILSTLILCSLCYGDMENTLFGAIRRDCULLRLVDDFLVTPHULTH	1154
OY	842	IRGSYVQOCGIPGSGILSTLILCSLCYGDMENTLFGAIRRDCULLRLVDDFLVTPHULTH	901
Db	1155	AKTFLRLVNGVPEYGCYVNLKRTVYVNPVEDALGCTAFVQMPAHGLFPWCGLLDTRT	1214
OY	902	AKTFLRLVNGVPEYGCYVNLKRTVYVNPVEDALGCTAFVQMPAHGLFPWCGLLDTRT	961
Db	1215	LEVOSDSSYARFSIRASVFNENGFKAQRMRKLELGVRLKCHSLFDLDQVNSIQYCT	1274
OY	962	LEVOSDSSYARFSIRASLTFNENGFKAQRMRKLELGVRLKCHSLFDLDQVNSIQYCT	1021
Db	1275	NIYKILLQAYRFHACVYLOLPHQOQWKNPTEFLRYISDTASLCSYILKAKNAGMSLAK	1334
OY	1022	NIYKILLQAYRFHACVYLOLPHQOQWKNPTEFLRYISDTASLCSYILKAKNAGMSLAK	1081
Db	1335	GAAGPLPSEAVOWMLCHQAFLLKTLRRRYTVVPLGSLRTAQOTQLSRKLPGTTLALEAA	1394
OY	1082	GAAGPLPSEAVOWMLCHQAFLLKTLRRRYTVVPLGSLRTAQOTQLSRKLPGTTLALEAA	1141
Db	1395	NPALPSPDKTILD 1407	
OY	1142	NPALPSPDKTILD 1154	
RESULT	12		
ID	US-08-974-549-628	STANDARD;	PRT; 1407 AA.
XX	xxxxxx		
DE	Sequence 628, Application US/08974549		
CC	GENERAL INFORMATION:		
CC	APPLICANT: Cecch, Thomas R.		
CC	APPLICANT: Lingner, Joachim		
CC	APPLICANT: Nakamura, Toru		
CC	APPLICANT: Chapman, Karen B.		
CC	APPLICANT: Morin, Gregg B.		
CC	APPLICANT: Harley, Calvin H.		
CC	APPLICANT: Andrews, William H.		
CC	TITLE OF INVENTION: Human Telomerase Catalytic Subunit		
CC	NUMBER OF SEQUENCES: 726		
CC	CORRESPONDENCE ADDRESS:		
CC	ADDRESSEE: Townsend and Townsend and Crew LLP		
CC	STREET: Two Embarcadero Center, Eighth Floor		
CC	CITY: San Francisco		
CC	STATE: California		
CC	COUNTRY: USA		
CC	ZIP: 94111-3834		
CC	COMPUTER READABLE FORM:		
CC	MEDIUM TYPE: Floppy disk		
CC	COMPUTER: IBM PC compatible		
CC	OPERATING SYSTEM: PC-DOS/MS-DOS		
CC	SOFTWARE: PatentIn Release #1.0, Version #1.30		
CC	CURRENT APPLICATION DATA:		

CC APPLICATION NUMBER: US/08/974,549  
CC FILING DATE: 19-NOV-1997  
CC CLASSIFICATION: 536  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/724,643  
CC FILING DATE: 01-OCT-1996  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/844,419  
CC FILING DATE: 18-APR-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/846,017  
CC FILING DATE: 25-APR-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/851,843  
CC FILING DATE: 06-MAY-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/854,050  
CC FILING DATE: 09-MAY-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/911,312  
CC FILING DATE: 14-AUG-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/912,951  
CC FILING DATE: 14-AUG-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/915,503  
CC FILING DATE: 14-AUG-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: WO PCT/US97/17618  
CC FILING DATE: 01-OCT-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: WO PCT/US97/17885  
CC FILING DATE: 01-OCT-1997  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Apple, Randolph Ted  
CC REGISTRATION NUMBER: 36,429  
CC REFERENCE/DOCKET NUMBER: 015389-002610US  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 576-0200  
CC TELEFAX: (415) 576-0300  
CC INFORMATION FOR SEQ ID NO: 628:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1407 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS:  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC FEATURE:  
CC NAME/KEY: Protein  
CC LOCATION: 1..1407  
CC OTHER INFORMATION: /note="fusion protein composed of  
CC OTHER INFORMATION: enhanced green fluorescent protein  
CC OTHER INFORMATION:  
CC OTHER INFORMATION: 5' untranslated region of hTERT mRNA and  
CC OTHER INFORMATION: hTERT protein sequence"  
CC SEQUENCE 1407 AA; 157668 MM; 10134798 CN;  
Query Match 99.6%; Score 8590; DB 18; Length 1407;  
Best Local Similarity 99.7%; Pred. No. 0.00e+00;  
Matches 1150; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Dh 255 ASTORCVLLRTMEALAPATPAMPRAPCRAVRSLLRSHYREVLPLATFPRLGPGGWRIV 314  
Dh 2 ASGQCVLLRTWEALAPATPAMPRAPCRAVRSLLRSHYREVLPLATFPRLGPGGWRIV 61  
Dh 315 QRGDPAARALVAOCLVCPMDARPPAPSPROYSCLELVARVQLRCEGAKNVIAF 374  
Dh 62 QRGDPAARALVAOCLVCPMDARPPAPSPROYSCLELVARVQLRCEGAKNVIAF 121  
Dh 375 GFALLDGAARGGPEAFVTSVRSYLPNTVTDALRGSGAMGLLRRVGDVVLHLARCALE 434  
Dh 122 GFALLDGAARGGPEAFVTSVRSYLPNTVTDALRGSGAMGLLRRVGDVVLHLARCALE 181

Dh 435 VLVAPSCAYQYCGPPLVYOLGATQATAPPPHASGPRRRLGCERAMNHSVEAGVPLGLPAP 494  
Dh 132 VLVAPSCAYQYCGPPLVYOLGATQATAPPPHASGPRRRLGCERAMNHSVEAGVPLGLPAP 241  
Dh 495 GARRRGSGASSSLPLPKRRRGGAPEPERTVGGGSMWHPGRTGPSDRGFCVSPAPPA 554  
Dh 242 GARRRGSGASSSLPLPKRRRGGAPEPERTVGGGSMWHPGRTGPSDRGFCVSPAPPA 301  
Dh 555 BEATSLGALSGTRHSHSPVGRHAGPSTSRPPMDTDCPPYATATKHEFLVSSGDK 614  
Dh 302 BEATSLGALSGTRHSHSPVGRHAGPSTSRPPMDTDCPPYATATKHEFLVSSGDK 361  
Dh 615 QLRPSFLSLRPSLTGARRLVETIFLGSRRPMGTPRRLRLPQRYQMPLFELILGN 674  
Dh 362 QLRPSFLSLRPSLTGARRLVETIFLGSRRPMGTPRRLRLPQRYQMPLFELILGN 421  
Dh 675 HAQCPYGVLLTCHPCLRAAVTPAAGVCAREKPGGSVAAPBEDDPRRLVOLLROHSSPW 734  
Dh 422 HAQCPYGVLLTCHPCLRAAVTPAAGVCAREKPGGSVAAPBEDDPRRLVOLLROHSSPW 481  
Dh 735 QYGVGRACLRVLPVPGIWSGRHNERRLRTKTKFISLGRKAKLSLOELTKMWSVROCAW 794  
Dh 482 QYGVGRACLRVLPVPGIWSGRHNERRLRTKTKFISLGRKAKLSLOELTKMWSVROCAW 541  
Dh 795 LRSRPGVCVPAAEHRLKEELIAFLHMLMSVYVELLSRFEYVETTFQKNRLFEYRPS 854  
Dh 542 LRSRPGVCVPAAEHRLKEELIAFLHMLMSVYVELLSRFEYVETTFQKNRLFEYRPS 601  
Dh 855 VWSKLQSIGIHOHLKRVOLRELSEAEVQRHREARPALITSRLRFPKPDGLRPVNDYV 914  
Dh 602 VWSKLQSIGIHOHLKRVOLRELSEAEVQRHREARPALITSRLRFPKPDGLRPVNDYV 661  
Dh 915 VGARFREREKRAERLTSVKALFVLANERARRPGLLGASVIGDDIHARARTFVLRRRA 974  
Dh 662 VGARFREREKRAERLTSVKALFVLANERARRPGLLGASVIGDDIHARARTFVLRRRA 721  
Dh 975 QDPPELTFVAVDTGADTIPODRLEVIASIIKPONTYCVRRAYVQKAHGHVRAAF 1034  
Dh 722 QDPPELTFVAVDTGADTIPODRLEVIASIIKPONTYCVRRAYVQKAHGHVRAAF 781  
Dh 1035 KSHVSTLDLPYMRQFVAHLQETSPRLDAVYIEOSSLSNASSGLEDFVFLRPMCHNAVR 1094  
Dh 782 KSHVSTLDLPYMRQFVAHLQETSPRLDAVYIEOSSLSNASSGLEDFVFLRPMCHNAVR 841  
Dh 1095 IRGKSYVOCGIPQGSILSTLCSLCYGDMDENKLEFAGIRBGLLRLVDDPLVTPHLTH 1154  
Dh 842 IRGKSYVOCGIPQGSILSTLCSLCYGDMDENKLEFAGIRBGLLRLVDDPLVTPHLTH 901  
Dh 1155 AKTEFLRTLVRGVPEYGCYVNLKRTVYVNFVEDEALGTAFFQMPAHGLFPMCGLLDTRT 1214  
Dh 902 AKTEFLRTLVRGVPEYGCYVNLKRTVYVNFVEDEALGTAFFQMPAHGLFPMCGLLDTRT 961  
Dh 1215 LEVQSDYSYVARTSIRASVTENRGFKAGRNMRKLLFGVLRKCHSLFDLDVNSLQVCT 1274  
Dh 962 LEVQSDYSYVARTSIRASVTENRGFKAGRNMRKLLFGVLRKCHSLFDLDVNSLQVCT 1021  
Dh 1275 NIYKTLILQAFRHACVQLQTFHQQVWKNPFEFLRVISDTASISCTSIKANAAGSGAK 1334  
Dh 1022 NIYKTLILQAFRHACVQLQTFHQQVWKNPFEFLRVISDTASISCTSIKANAAGSGAK 1081  
Dh 1335 GAAGPLPSEAVOMLCHQAFLLKTRHRTYVPLLSLRTAOTLSRKLPGTTILALEANA 1394  
Dh 1082 GAAGPLPSEAVOMLCHQAFLLKTRHRTYVPLLSLRTAOTLSRKLPGTTILALEANA 1141  
Dh 1395 NPALPSEDKTILD 1407  
Dh 1142 NPALPSEDKTILD 1154

RESULT 13  
ID US-08-912-951-334 STANDARD; PRT: 1407 AA.  
XX  
AC xxxxxx

XX XX  
 DE DE  
 CC Sequence 334, Application US/08912951  
 CC  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Cech, Thomas R.  
 CC APPLICANT: Lingner, Joachim  
 CC APPLICANT: Nakamura, Toru  
 CC APPLICANT: Chapman, Karen B.  
 CC APPLICANT: Morin, Gregg B.  
 CC APPLICANT: Harley, Calvin  
 CC APPLICANT: Andrews, William H.  
 CC TITLE OF INVENTION: HUMAN TELOMERASE CATALYTIC SUBUNIT: DIAGNOSTIC AND  
 CC THERAPEUTIC METHODS  
 CC NUMBER OF SEQUENCES: 335  
 CC CORRESPONDENCE ADDRESSES:  
 CC ADDRESSEE: Townsend and Townsend and Crew LLP  
 CC STREET: Two Embarcadero Center, 8th Floor  
 CC CITY: San Francisco  
 CC STATE: California  
 CC COUNTRY: United States of America  
 CC ZIP: 94111  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.30  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/912,951  
 CC FILING DATE: 14-AUG-1997  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/854,050  
 CC FILING DATE: 09-MAY-1997  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/846,017  
 CC FILING DATE: 06-MAY-1997  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/851,843  
 CC FILING DATE: 18-APR-1997  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/844,419  
 CC FILING DATE: 18-APR-1997  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/724,643  
 CC FILING DATE: 01-OCT-1996  
 CC CLASSIFICATION: 435  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Apple, Randolph T.  
 CC REGISTRATION NUMBER: 36,429  
 CC REFERENCE/DOCKET NUMBER: 015389-00260005  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (415) 576-0200  
 CC TELEFAX: (415) 576-0300  
 CC INFORMATION FOR SEQ ID NO: 334:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 1407 amino acids  
 CC TYPE: amino acid  
 CC STRANDEDNESS:  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 1407 AA; 157668 MW; 10134798 CN;  
 CC  
 CC Query Match 99.6%; Score 8590; DB 13; Length 1407;  
 CC Best Local Similarity 99.7%; Pred. No. 0.00e+00;  
 CC Matches 1150; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 255 ASTORCVLLRTWEALAPATPAMPAPRCRAVRSLSRSHREVLPPLATEFVRRLGPOGRWLV 314  
 QY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 2 ASGORCVLLRTWEALAPATPAMPAPRCRAVRSLSRSHREVLPPLATEFVRRLGPOGRWLV 61  
 Db 315 QRGDPAAFRALVAOCLVCPWDAPPPAPPSFROYSCLELVARYLQRCERGAKNVAF 374  
 QY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 62 QRGDPAAFRALVAOCLVCPWDAPPPAPPSFROYSCLELVARYLQRCERGAKNVAF 121  
 Db 375 GFALLDARGGPPPEAFSTSVRSYLPNTYTDALRGSGAGLLRRYGDVVLVHLARCLF 434  
 QY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 122 GFALLDARGGPPPEAFSTSVRSYLPNTYTDALRGSGAGLLRRYGDVVLVHLARCLF 181  
 Db 435 VLVAPSCAYOVCPPPLYOLGAATQAPPPHAGSPRRRLGCRAMNHSVREAGVPLGLPAP 494  
 QY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 182 VLVAPSCAYOVCPPPLYOLGAATQAPPPHAGSPRRRLGCRAMNHSVREAGVPLGLPAP 241  
 Db 495 GARRRGASASRLPLKPRRRGAAPPEPERTVVGOGSWAHPTRGPSDRGFCVSPAPPA 554  
 QY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 242 GARRRGASASRLPLKPRRRGAAPPEPERTVVGOGSWAHPTRGPSDRGFCVSPAPPA 301  
 Db 555 EEAATLEAGLSGRSHSPSVGRHAGPSTSRPPRPMDTCPPYATATKTFVLSGDK 614  
 QY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 302 EEAATLEAGLSGRSHSPSVGRHAGPSTSRPPRPMDTCPPYATATKTFVLSGDK 361  
 Db 615 QLRPSFLSLRPSLTGARLVETFTLGRPMMPGTPRRLPDRYQOMPLFLELGN 674  
 QY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 362 QLRPSFLSLRPSLTGARLVETFTLGRPMMPGTPRRLPDRYQOMPLFLELGN 421  
 Db 675 HAOCYVGLKTHCPRLRAAVTPAAGVCAREKPOGSVAAPBEEDTPRRLVOLLROHSSPW 734  
 QY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 422 HAOCYVGLKTHCPRLRAAVTPAAGVCAREKPOGSVAAPBEEDTPRRLVOLLROHSSPW 481  
 Db 735 QYGGVRACTRLVPPGGLMSRHNERRLRNTKFKISLQKNAKLSQDLTKMSVRDCAW 794  
 QY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 482 QYGGVRACTRLVPPGGLMSRHNERRLRNTKFKISLQKNAKLSQDLTKMSVRDCAW 541  
 Db 795 LRRSPGVCVPAAEHRLREELIAKFLHMLMSYVVELRSFFVYETETFOKNRLEFPSP 854  
 QY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 542 LRRSPGVCVPAAEHRLREELIAKFLHMLMSYVVELRSFFVYETETFOKNRLEFPSP 601  
 Db 855 VWSKLQSIGIRQHLKRVOLRELSEAEVROHREARPAALLTSRLRFPKPDGLRPIYMDYV 914  
 QY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 602 VWSKLQSIGIRQHLKRVOLRELSEAEVROHREARPAALLTSRLRFPKPDGLRPIYMDYV 661  
 Db 915 VGARTFRREKRAERLTSVKALFSYLANERARPCGLCAVGLDDHRAKTEFVLARVA 974  
 QY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 662 VGARTFRREKRAERLTSVKALFSYLANERARPCGLCAVGLDDHRAKTEFVLARVA 721  
 Db 975 QDPPELTFVAVDVTGAYDTLPDRLTEVIASITKPONTYCVRRAYVQKAHGHVRAAF 1034  
 QY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 722 QDPPELTFVAVDVTGAYDTLPDRLTEVIASITKPONTYCVRRAYVQKAHGHVRAAF 781  
 Db 1035 KSHVSTLTDLPYMRFOFVAHLQETSPLRDAAVIEQSSSLNEASSGLFDFVLRFCMHAAVR 1094  
 QY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 782 KSHVSTLTDLPYMRFOFVAHLQETSPLRDAAVIEQSSSLNEASSGLFDFVLRFCMHAAVR 841  
 Db 1095 IRGKSYVOCGIPQGSILSTLCSLCYDMDENKLFAGIRDGLLRVLDVDFLVTPLHLTH 1154  
 QY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 842 IRGKSYVOCGIPQGSILSTLCSLCYDMDENKLFAGIRDGLLRVLDVDFLVTPLHLTH 901  
 Db 1155 AKTEFLRTLVRCVPEYGCYVNNLRKTYVNFVDEALGTAFAVQMPAHGLFPMCGLLDRT 1214  
 QY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 902 AKTEFLRTLVRCVPEYGCYVNNLRKTYVNFVDEALGTAFAVQMPAHGLFPMCGLLDRT 961  
 Db 1215 LEVQSDYSYARTSIRASVTFRNGFKAGRNMRKLFGLVLRKCHSLFIDLQVNSIQTYCT 1274  
 QY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 962 LEVQSDYSYARTSIRASVTFRNGFKAGRNMRKLFGLVLRKCHSLFIDLQVNSIQTYCT 1021  
 Db 1275 NIYKTLILQATRFACVATQLPFRHOQWKNPFFELFVVISDTSLSCTIILKAKNAGSLAK 1334  
 QY ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 1022 NIYKTLILQATRFACVATQLPFRHOQWKNPFFELFVVISDTSLSCTIILKAKNAGSLAK 1081  
 Db 1335 GAAGPLPSEAVQWLCHQAFLLKLTGRHRTVYVPLGSLTAQTQSRKLRGTTTLALEAA 1394

|||||  
QY 1082 GAAPLSEAVQWCHQAFILKTRHRTYTPPLSGLSFTAQOTLSRKLPGTTLALEAA 1141  
Db 1395 NPALPSDEKTLID 1407  
QY 1142 NPALPSDEKTLID 1154  
RESULT 14  
ID US-08-974-549-2 STANDARD: PRT: 1132 AA.  
AC xxxxxx  
XX  
XX  
XX  
DE Sequence 2, Application US/08974549  
XX  
CC Sequence 2, Application US/08974549  
CC GENERAL INFORMATION:  
CC APPLICANT: Cech, Thomas R.  
CC APPLICANT: Lingner, Joachim  
CC APPLICANT: Nakamura, Toru  
CC APPLICANT: Chapman, Karen B.  
CC APPLICANT: Morin, Gregg B.  
CC APPLICANT: Harley, Calvin B.  
CC APPLICANT: Andrews, William H.  
CC TITLE OF INVENTION: Human Telomerase Catalytic Subunit  
CC NUMBER OF SEQUENCES: 726  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Townsend and Townsend and Crew LLP  
CC STREET: Two Embarcadero Center, Eighth Floor  
CC City: San Francisco  
CC STATE: California  
CC COUNTRY: USA  
CC ZIP: 94111-3834  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/974,549  
CC FILING DATE: 19-NOV-1997  
CC CLASSIFICATION: 536  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/724,643  
CC FILING DATE: 01-OCT-1996  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/844,419  
CC FILING DATE: 18-APR-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/846,017  
CC FILING DATE: 25-APR-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/851,843  
CC FILING DATE: 06-MAY-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/854,050  
CC FILING DATE: 09-MAY-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/911,312  
CC FILING DATE: 14-AUG-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/912,951  
CC FILING DATE: 14-AUG-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/915,503  
CC FILING DATE: 14-AUG-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: WO PCT/US97/17618  
CC FILING DATE: 01-OCT-1997  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: WO PCT/US97/17885

CC FILING DATE: 01-OCT-1997  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Apple, Randolph Ted  
CC REGISTRATION NUMBER: 36,429  
CC REFERENCE/DOCKET NUMBER: 015389-0026100S  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 576-0200  
CC TELEFAX: (415) 576-0300  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1132 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 1132 AA, 126995 MW, 6588209 CN;  
SQ  
Query Match 98.2%; Score 8465; DB 18; Length 1132;  
Best Local Similarity 100.0%; Pred. No. 0.00e+00;  
Matches 1132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 1 MPAPRCRAVRSLSLRSHRYEVLPATVFRRLGPOGWRVLVORGPAPAFALVAQCIVCPW 60  
QY 23 MPAPRCRAVRSLSLRSHRYEVLPATVFRRLGPOGWRVLVORGPAPAFALVAQCIVCPW 82  
Db 61 DARPPAPSPFROVSCLELVARVLORLCERGAKNVLAFFALDGAAGGPPFAFTTSVR 120  
QY 83 DARPPAPSPFROVSCLELVARVLORLCERGAKNVLAFFALDGAAGGPPFAFTTSVR 142  
Db 121 SYLPNTVTDALRGSGAGLLRLRYGDDVLVHLARCALFVLVAPSCAYQVCGPPLYQLGA 180  
QY 143 SYLPNTVTDALRGSGAGLLRLRYGDDVLVHLARCALFVLVAPSCAYQVCGPPLYQLGA 202  
Db 181 ATQARPPPHASGPPRRRLGGERANNSYREAGVPLGLPAPARRRGSGASSLPLPKPRR 240  
QY 203 ATQARPPPHASGPPRRRLGGERANNSYREAGVPLGLPAPARRRGSGASSLPLPKPRR 262  
Db 241 GAAPERTPVGQSWAHPPRTGSPDRGFCVVSPPAPAEATSLGALGTRSHPSVG 300  
QY 263 GAAPERTPVGQSWAHPPRTGSPDRGFCVVSPPAPAEATSLGALGTRSHPSVG 322  
Db 301 ROHHAGPSTSRPPMDTCCPPYATKHFLLYSSGDEQLRPSFLLSLRPSLTGARL 360  
QY 323 ROHHAGPSTSRPPMDTCCPPYATKHFLLYSSGDEQLRPSFLLSLRPSLTGARL 382  
Db 361 VETFFLSRPMWGPTRRLRPLPORTWQMRPLLELIGNAOCGYVLLTTHCPRLRAVT 420  
QY 383 VETFFLSRPMWGPTRRLRPLPORTWQMRPLLELIGNAOCGYVLLTTHCPRLRAVT 442  
Db 421 PAAGVCAREKPOGSVAAPBEDDPRRLVOLLROHSSPMQVVGVRACLRRLVPPGLMGS 480  
QY 443 PAAGVCAREKPOGSVAAPBEDDPRRLVOLLROHSSPMQVVGVRACLRRLVPPGLMGS 502  
Db 481 RHNERRLRMTKKEISLGKHAKLISLOELTWKMSYRDCAMLRSPGVGCVPAAEHRLREEI 540  
QY 503 RHNERRLRMTKKEISLGKHAKLISLOELTWKMSYRDCAMLRSPGVGCVPAAEHRLREEI 562  
Db 541 LAKFLHMLSYVYVELLRFFTYTETTFQKNRLEFFTKSWSKLSQISGIOHLKRVOLRE 600  
QY 563 LAKFLHMLSYVYVELLRFFTYTETTFQKNRLEFFTKSWSKLSQISGIOHLKRVOLRE 622  
Db 601 LSEAEVQOHEARPALTLTSRLRFPKPDGLRPIYNMDDYVGARFRRERKAERLTSVKA 660  
QY 623 LSEAEVQOHEARPALTLTSRLRFPKPDGLRPIYNMDDYVGARFRRERKAERLTSVKA 682  
Db 661 LFSVLYNTERARRPGLGASVYLGDDIHRAMRTVYLRYRAODPPPELTFVAVDTYGAADTI 720  
QY 683 LFSVLYNTERARRPGLGASVYLGDDIHRAMRTVYLRYRAODPPPELTFVAVDTYGAADTI 742  
Db 721 PODRLTEVIASIIKPPONTYCVRRYAVVQKAAGHVRAKAFESHVSTLTDLPYRQFVAHL 780  
QY 743 PODRLTEVIASIIKPPONTYCVRRYAVVQKAAGHVRAKAFESHVSTLTDLPYRQFVAHL 802  
Db 781 QETSPFLDAVAVIEGSSSLNASSGLFDVFLRFPMCHAHAVRLRGKSYVQCGGIIPOGSILSTL 840

|||||  
 QY 803 QETSPLRDAVIEOSSSLNEASSGLFEDVFLRFCHAVIRKSYQCCGIPQGSILSTL 862  
 Db 841 LCLSCVGDMENTLFAIRIRBDGLLRVDVDFLVTPLHAKTFLRLVNGVPEYGVNL 900  
 QY 863 LCLSCVGDMENTLFAIRIRBDGLLRVDVDFLVTPLHAKTFLRLVNGVPEYGVNL 922  
 Db 901 RTVYVPEDEALGTAFAVQMPAHGLFPMCGLLDTRLEVOSSYSYARTISASLTF 960  
 QY 923 RTVYVPEDEALGTAFAVQMPAHGLFPMCGLLDTRLEVOSSYSYARTISASLTF 982  
 Db 961 NRGFKAGMRRLKFGVLRKCHSLFLDQVNSLQVCTNRYKILLQAVRHFACVQLP 1020  
 QY 983 NRGFKAGMRRLKFGVLRKCHSLFLDQVNSLQVCTNRYKILLQAVRHFACVQLP 1042  
 Db 1021 FHOQVKNFTFLRVISDPSLCLYSLKAKNAGMSLGAGAGAPLSEAVQVLCQAFL 1080  
 QY 1043 FHOQVKNFTFLRVISDPSLCLYSLKAKNAGMSLGAGAGAPLSEAVQVLCQAFL 1102  
 Db 1081 KLTRHRTVYVPLGLSLRTAOTOLSRKLPGLTTLFALPAAAPALPSDFKTLID 1132  
 QY 1103 KLTRHRTVYVPLGLSLRTAOTOLSRKLPGLTTLFALPAAAPALPSDFKTLID 1154  
 RESULT 15  
 ID US-08-912-951-2 STANDARD: PRF: 1132 AA.  
 AC xxxxxx  
 DT  
 XX  
 XX  
 DE Sequence 2, Application US/08912951  
 CC  
 CC Sequence 2, Application US/08912951  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Cecch, Thomas R.  
 CC APPLICANT: Lingner, Joachim  
 CC APPLICANT: Nakamura, Toru  
 CC APPLICANT: Chapman, Karen B.  
 CC APPLICANT: Morin, Gregg B.  
 CC APPLICANT: Harley, Calvin  
 CC APPLICANT: Andrews, William H.  
 CC TITLE OF INVENTION: HUMAN TELOMERASE CATALYTIC SUBUNIT: DIAGNOSTIC AND  
 CC TITLE OF INVENTION: THERAPEUTIC METHODS  
 CC NUMBER OF SEQUENCES: 335  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESS: Townsend and Townsend and Crew LLP  
 CC STREET: Two Embarcadero Center, 8th floor  
 CC CITY: San Francisco  
 CC STATE: California  
 CC COUNTRY: United States of America  
 CC ZIP: 94111  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patent Release #1.0, Version #1.30  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/912,951  
 CC FILING DATE: 14-AUG-1997  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/854,050  
 CC FILING DATE: 09-MAY-1997  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/851,843  
 CC FILING DATE: 06-MAY-1997  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/846,017  
 CC FILING DATE: 25-APR-1997  
 CC CLASSIFICATION: 435

CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/844,419  
 CC FILING DATE: 18-APR-1997  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/724,643  
 CC FILING DATE: 01-OCT-1996  
 CC CLASSIFICATION: 435  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Apple, Randolph T.  
 CC REGISTRATION NUMBER: 36,429  
 CC REFERENCE/DOCKET NUMBER: 015389-002600US  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (415) 576-0200  
 CC TELEFAX: (415) 576-0300  
 CC INFORMATION FOR SEQ ID NO: 2:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 1132 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 1132 AA; 12695 MW; 6588209 CN;  
 SQ  
 Query Match 98.2%; Score 8465; DB 13; Length 1132;  
 Best Local Similarity 100.0%; Pred. No. 0.00e+00;  
 Matches 1132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 1 MPAPRCRAVRSLSHREVLPLATFVRRLGPGQMRVQGDPAFALVACIVCPW 60  
 QY 23 MPAPRCRAVRSLSHREVLPLATFVRRLGPGQMRVQGDPAFALVACIVCPW 82  
 Db 61 DARPPAPASFROVSGCKELVAVLQRCERGAKNVLAFGFALDGAAGGPEAFTTSR 120  
 QY 83 DARPPAPASFROVSGCKELVAVLQRCERGAKNVLAFGFALDGAAGGPEAFTTSR 142  
 Db 121 SYLPTVTYDALRGSGAMGILLRRVGDVYLHLARCAFLVAPSCAYOVCPPIYOLGA 180  
 QY 143 SYLPTVTYDALRGSGAMGILLRRVGDVYLHLARCAFLVAPSCAYOVCPPIYOLGA 202  
 Db 181 AFOARPPHAGSPRRRLGGERAMNHSVREAGVPLGLPAPGARRGSGASRSJPLPKRPR 240  
 QY 203 AFOARPPHAGSPRRRLGGERAMNHSVREAGVPLGLPAPGARRGSGASRSJPLPKRPR 262  
 Db 241 GAAPERTPVGQSWAHGRTGRGSDRGFCVSPAPAPAEATSLGALSGRHSHPVG 300  
 QY 263 GAAPERTPVGQSWAHGRTGRGSDRGFCVSPAPAPAEATSLGALSGRHSHPVG 322  
 Db 301 RQHHAGPSTSPRPWDTPCPVYAEIKHFLYSSGDKQOLRPSFLSLRSLTGARL 360  
 QY 323 RQHHAGPSTSPRPWDTPCPVYAEIKHFLYSSGDKQOLRPSFLSLRSLTGARL 382  
 Db 361 VETIFLGSPPMPGTPRRRLPQRYWQMRPLFELLGNAHCOPYGVILKTCPLPAVT 420  
 QY 383 VETIFLGSPPMPGTPRRRLPQRYWQMRPLFELLGNAHCOPYGVILKTCPLPAVT 442  
 Db 421 PAAGCAREKPGSVAAPEEDTPRRVLQRLROHSSPMQVGVFACLRRLVPPGINGS 480  
 QY 443 PAAGCAREKPGSVAAPEEDTPRRVLQRLROHSSPMQVGVFACLRRLVPPGINGS 502  
 Db 481 RNNERFLRNTKFTSLGKHAFLSLQELTWKMSVDCAMLRSPVGCVPAAEHLREI 540  
 QY 503 RNNERFLRNTKFTSLGKHAFLSLQELTWKMSVDCAMLRSPVGCVPAAEHLREI 562  
 Db 541 LAKFLHMLSVYVELLSFFYVETETQKNRLEFFRYSWSKLOSIGIRQLKVOURE 600  
 QY 563 LAKFLHMLSVYVELLSFFYVETETQKNRLEFFRYSWSKLOSIGIRQLKVOURE 622  
 Db 601 LSEAVRQRRARPALTLRLRIFIKPDGLRPIYMDVYVAGRTRRERARLRSLRYA 660  
 QY 623 LSEAVRQRRARPALTLRLRIFIKPDGLRPIYMDVYVAGRTRRERARLRSLRYA 682  
 Db 661 LPSVLYERARRPGLLGASVGLDDIHRAMRTFVLRVRAQDPPPLFYKVDVTGAYTI 720

```

Oy 683 LFSVLNTERARRPGLIGASVLGLDDIHRAMTFVLNRADPEPELYFVKVDVTGAYDTI 742
Db 721 PODRLTEVIASIIKPONTYCVARRAVVOKAAAGHVKAFAKSHVSTLTDLOPYMRQFVAHL 780
Oy 743 PODRLTEVIASIIKPONTYCVARRAVVOKAAAGHVKAFAKSHVSTLTDLOPYMRQFVAHL 802
Db 781 OETSPLRDAVVIQSSSLNEASSGLFDVFLRPMCHHAVIRGKSYVOCGIPQGSILSTL 840
Oy 803 OETSPLRDAVVIQSSSLNEASSGLFDVFLRPMCHHAVIRGKSYVOCGIPQGSILSTL 862
Db 841 LCSLCYGDMEKFLFAGIRRDGILLRLVDDFLLVTPHLTHAKTFLRTLVRGPEYGCYVNL 900
Oy 863 LCSLCYGDMEKFLFAGIRRDGILLRLVDDFLLVTPHLTHAKTFLRTLVRGPEYGCYVNL 922
Db 901 RRTVYNFPEDEALGTAFAVQMPAHGLFPWCGLLDDRTLLEYQSDYSSYARTSIRASLTF 960
Oy 923 RRTVYNFPEDEALGTAFAVQMPAHGLFPWCGLLDDRTLLEYQSDYSSYARTSIRASLTF 982
Db 961 NRGFRAGRMRRKFLFVLRKLKCHSLFLDLQVNSIQTVCTNIYKILLQAYRPHACVLOLP 1020
Oy 983 NRGFRAGRMRRKFLFVLRKLKCHSLFLDLQVNSIQTVCTNIYKILLQAYRPHACVLOLP 1042
Db 1021 FHOQVWKNPTEFLRVISDTASLCYSILKAKNAGMSLGAKGAGPLPSEAVQWLCQAFL 1080
Oy 1043 FHOQVWKNPTEFLRVISDTASLCYSILKAKNAGMSLGAKGAGPLPSEAVQWLCQAFL 1102
Db 1081 KTRRRVTVVPLLGSLRTQOTLSRKLPGTTTALAEAAANPALPSDEKTIID 1132
Oy 1103 KTRRRVTVVPLLGSLRTQOTLSRKLPGTTTALAEAAANPALPSDEKTIID 1154

```

Search completed: Fri Dec 18 18:34:40 1998  
 Job time : 197 secs.



```

CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER:  US 621, 849
CC      FILING DATE:  04-DEC-1990
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER:  US 616, 374
CC      FILING DATE:  21-NOV-1990
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER:  US 600, 024
CC      FILING DATE:  18-OCT-1990
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER:  US 599, 543
CC      FILING DATE:  18-OCT-1990
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER:  US 579, 865
CC      FILING DATE:  07-SEP-1990
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER:  US 569, 920
CC      FILING DATE:  20-AUG-1990
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER:  US 483, 913
CC      FILING DATE:  22-FEB-1990
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER:  US 422, 613
CC      FILING DATE:  17-OCT-1989
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER:  US 315, 342
CC      FILING DATE:  23-FEB-1989
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER:  US 232, 630
CC      FILING DATE:  15-AUG-1988
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER:  US 179, 460
CC      FILING DATE:  08-APR-1988
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME:  PITCHER, EDMUND R.
CC      REGISTRATION NUMBER:  27, 829
CC      REFERENCE/DOCKET NUMBER:  CRP-001CP6
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE:  617/248-7000
CC      TELEFAX:  617/248-7100
CC      INFORMATION FOR SEQ ID NO:  29:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH:  402 amino acids
CC      TYPE:  AMINO ACID
CC      TOPOLOGY:  linear
CC      MOLECULE TYPE:  protein
CC      SEQUENCE  402 AA;  44698 MW;  800499 CN;

```

Query Match		1.8%	Score 125;	DB 1;	Length 402;
Best Local Similarity		39.7%;	Pred. No. 5.57e-01;		
Matches	25;	Conservative	12;	Mismatches 21;	Indels 5; Gaps 5;
Dd	12	GATCAGGCGGPRLPDP-GC-PORRLG-AEREDVOREILAVLGLPGR-PRRAPPAAS	67		
	:	:	:      :	:	:
Qy	194	GPRLYOIGAATQA-RPPPHASGPRRRLGCERAMNHVSREAGVPLGLTPAPGARRRGSASR	252		
Dd	68	RLP	70		
	:				
Qy	253	SLP	255		
RESULT	2				
ID	US-08-643-763A-21	STANDARD;	PRT;	402	AA.
XX					
AC	xxxxxx				
XX					
DT					
XX					
DE	Sequence 21, Application US/08643763A				
CC	Sequence 21, Application US/08643763A				
CC	Patent No. 5733878				
CC	GENERAL INFORMATION:				
CC					

CC APPLICANT: KUBERSAMPATH, THANQAVEL  
 CC APPLICANT: RUGGER, DAVID C.  
 CC APPLICANT: OPPERMANN, HERMAN  
 CC APPLICANT: COHEN, CHARLES M.  
 CC APPLICANT: PANG, ROY H. L.  
 CC TITLE OF INVENTION: MORPHOGENIC-INDUCED PERIODONTAL TISSUE  
 CC TITLE OF INVENTION: REGENERATION.  
 CC NUMBER OF SEQUENCES: 33  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES  
 CC STREET: 45 SOUTH STREET  
 CC CITY: HOPKINTON  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 01748  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/643,763A  
 CC FILING DATE: 06-MAY-1996  
 CC CLASSIFICATION: 514  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: FENTON ESQ., GILLIAN M.  
 CC REGISTRATION NUMBER: 36,508  
 CC REFERENCE/DOCKET NUMBER: CRP-06/CN  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 248-7560  
 CC TELEFAX: (617) 248-7100  
 CC INFORMATION FOR SEQ ID NO: 21:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 402 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 402 AA; 44764 MW; 803627 CN;

Query Match	1.8%	Score 125;	DB 1;	Length 402;
Best Local Similarity	39.7%	Pred. No. 5.57e-01;		
Matches 25;	Conservative	12;	Mismatches 21;	Indels 5; Gaps 5
Db	12	GLALCALGGGPGRLRPP-GC-PQRILG-ARERDVOREILAVLGR-PRRAPPAAS	67	
QY	194	GPPLYLQGAANYA-RPPPHASGPRRLRGCCERAMNHVSREAGVPLGLDPAFGARRRGSASR	252	
Db	68	RLP 70		
QY	253	SLP 255		
RESULT 3				
ID	US-08-451-953A-21	STANDARD;	PRT;	402 AA.
XX	xxxxxx			
XX				
DT				
XX				
DE	Sequence 21, Application US/08451953A			
XX				
CC	Sequence 21, Application US/08451953A			
CC	Patent No. 5741641			
CC	GENERAL INFORMATION:			
CC	APPLICANT: SMART, JOHN			
CC	APPLICANT: OPPERMAN, HERMAN			
CC	APPLICANT: OZKAYNAK, ENGIN			
CC	APPLICANT: KUBERASAMPATH, THANGAVEL			
CC	APPLICANT: RUEGER, DAVID C.			
CC	APPLICANT: PANG, ROY H. L.			
CC	APPLICANT: COHEN, CHARLES M.			
CC	TITLE OF INVENTION: MORPHOGENIC PROTEIN SCREENING METHOD			
CC	NUMBER OF SEQUENCES: 33			



[illegible]

CC	ATTORNEY/AGENT INFORMATION:
CC	NAME: PITCHER, EDMUND R.
CC	REGISTRATION NUMBER: 27,829
CC	REFERENCE/DOCKET NUMBER: CRP-096
CC	TELECOMMUNICATION INFORMATION:
CC	TELEPHONE: 617/7000
CC	TELEFAX: 617/248-7100
CC	INFORMATION FOR SEQ. ID NO: 4:
CC	SEQUENCE CHARACTERISTICS:
CC	LENGTH: 402 amino acids
CC	TYPE: amino acid
CC	TOPOLOGY: linear
CC	MOLECULE TYPE: protein
CC	SEQUENCE 402 AA; 44698 MM; 800499 CN;
CC	
CC	Query Match 1.8%; Score 125; DB 1; Length 402;
CC	Best Local Similarity 39.7%; Pred. No. 5.57e-01;
CC	Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;
CC	
Db	12 GLALCALGGSGPGRLRPP-GC-PQRRLG-ARERDVORELIANVLGLGR-PRRAPPAS 67
QY	194 GPPYLQGAATQA-RPPHASGPRRLRGCCERAMNHSVREAGVPLGLPAPGARRRGGSASR 252
Db	68 RLP 70
QY	253 SLP 255
CC	
RESULT	7 STANDARD; PRT; 402 AA.
ID	US-08-155-343A-21
XX	xxxxxx
DT	
XX	
DE	Sequence 21, Application US/08155343A
XX	
CC	Sequence 21, Application US/08155343A
CC	Patent No. 5656593
CC	GENERAL INFORMATION:
CC	APPLICANT: KUBERASAMPATH, THANGAVEL
CC	APPLICANT: RUEGER, DAVID C.
CC	APPLICANT: OPPERMANN, HERMAN
CC	APPLICANT: COHEN, CHARLES M.
CC	APPLICANT: PANG, ROY H. L.
CC	TITLE OF INVENTION: MORPHOGENIC-INDUCED PERIODONTAL TISSUE
CC	TITLE OF INVENTION: REGENERATION.
CC	NUMBER OF SEQUENCES: 33
CC	CORRESPONDENCE ADDRESS:
CC	ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
CC	STREET: 45 SOUTH STREET
CC	CITY: HOPKINTON
CC	STATE: MA
CC	COUNTRY: USA
CC	ZIP: 01748
CC	COMPUTER READABLE FORM:
CC	MEDIUM TYPE: floppy disk
CC	COMPUTER: IBM PC compatible
CC	OPERATING SYSTEM: PC-DOS/MS-DOS
CC	SOFTWARE: PatentIn Release #1.0, Version #1.30
CC	CURRENT APPLICATION DATA:
CC	APPLICATION NUMBER: US/08/155,343A
CC	FILING DATE: 15-NOV-1993
CC	CLASSIFICATION: 435
CC	ATTORNEY/AGENT INFORMATION:
CC	NAME: FENTON Esq., GILLIAN M.
CC	REGISTRATION NUMBER: 36,508
CC	REFERENCE/DOCKET NUMBER: CRP-067EW
CC	TELECOMMUNICATION INFORMATION:
CC	TELEPHONE: (617) 248-7560
CC	TELEFAX: (617) 248-7100
CC	INFORMATION FOR SEQ. ID NO: 21:
CC	SEQUENCE CHARACTERISTICS:
CC	

CC LENGTH: 402 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
SQ SEQUENCE 402 AA; 44764 MW; 803627 CN;

Query Match 1.8%; Score 125; DB 1; Length 402;  
Best Local Similarity 39.7%; Pred. No. 5.57e-01;  
Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;

Db 12 GLALCALGGGPGRLRPP-GC-PORRLG-ARERDVOREILAVGLPGR-PRPAPPAAS 67  
Y 194 GPPLYQLGAATQA-RPPHASGPRRLGCERAMNHSVREAGVPLGLPAPGARRRGGSASR 252

Db 68 RLP 70  
Y 253 SLP 255

RESULT 8  
ID PCT-US93-07231-21 STANDARD; PRT; 402 AA.  
XX AC xxxxxx

Sequence 21, Application PC/TUS9307231  
Sequence 21, Application PC/TUS9307231  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: MORPHOGEN-INDUCED NERVE REGENERATION AND  
TITLE OF INVENTION: REPAIR  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CREATIVE BIOMOLECULES, INC.  
STREET: 35 SOUTH STREET  
CITY: HOPKINTON  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 01748  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/07231  
FILING DATE: 19930729  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: KELLEY, ROBIN D.  
REGISTRATION NUMBER: 34,637  
REFERENCE/DOCKET NUMBER: CRP-070  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/248-7000  
TELEFAX: 617/248-7100  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 402 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE 402 AA; 44698 MW; 800499 CN;

Query Match 1.8%; Score 125; DB 2; Length 402;  
Best Local Similarity 39.7%; Pred. No. 5.57e-01;  
Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;

Db 12 GLALCALGGGPGRLRPP-GC-PORRLG-ARERDVOREILAVGLPGR-PRPAPPAAS 67  
Y 194 GPPLYQLGAATQA-RPPHASGPRRLGCERAMNHSVREAGVPLGLPAPGARRRGGSASR 252

Db 68 RLP 70  
Y 253 SLP 255

RESULT 9  
ID US-08-480-528A-8 STANDARD; PRT; 402 AA.  
XX AC xxxxxx

Sequence 8, Application US/08480528A  
Sequence 8, Application US/08480528A  
Patent No. 5652118  
GENERAL INFORMATION:  
APPLICANT: OPPERMAN, HERMANN  
APPLICANT: OZKAYNAK, ENGIN  
APPLICANT: KUBERASAMPATH, THANGAVEL  
APPLICANT: RUBGER, DAVID C.  
APPLICANT: PANG, ROY H.L.  
TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES  
STREET: 45 SOUTH STREET  
CITY: HOPKINTON  
STATE: MA  
COUNTRY: USA  
ZIP: 01748  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,528A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: FENTON Esq., GILLIAN M.  
REGISTRATION NUMBER: 36,508  
REFERENCE/DOCKET NUMBER: CRP-076FW  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7560  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 402 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE 402 AA; 44764 MW; 803627 CN;

Query Match 1.8%; Score 125; DB 1; Length 402;  
Best Local Similarity 39.7%; Pred. No. 5.57e-01;  
Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;

Db 12 GLALCALGGGPGRLRPP-GC-PORRLG-ARERDVOREILAVGLPGR-PRPAPPAAS 67  
Y 194 GPPLYQLGAATQA-RPPHASGPRRLGCERAMNHSVREAGVPLGLPAPGARRRGGSASR 252

Db 68 RLP 70  
Y 253 SLP 255

RESULT 10  
ID US-07-901-703-11 STANDARD; PRT; 402 AA.  
XX AC xxxxxx

```

XX DT
XX DE
XX Sequence 11, Application US/07901703
XX CC
XX Patent No. 5344654
XX CC
XX GENERAL INFORMATION:
XX CC
XX APPLICANT: RUEGER, DAVID C
XX CC
XX APPLICANT: KUBERASAMPATH, THANGAVEL
XX CC
XX APPLICANT: OPPERMANN, HERMANN
XX CC
XX TITLE OF INVENTION: PROSTHETIC DEVICES HAVING ENHANCED
XX CC
XX TITLE OF INVENTION: OSTEOGENIC PROPERTIES
XX CC
XX NUMBER OF SEQUENCES: 22
XX CC
XX CORRESPONDENCE ADDRESS:
XX CC
XX ADDRESSEE: TESTA, HURWITZ & THIBEAULT
XX CC
XX STREET: EXCHANGE PLACE, 53 STATE STREET
XX CC
XX CITY: BOSTON
XX CC
XX STATE: MA
XX CC
XX COUNTRY: USA
XX CC
XX ZIP: 02109
XX CC
XX COMPUTER READABLE FORM:
XX CC
XX MEDIUM TYPE: Floppy-disk
XX CC
XX COMPUTER: IBM PC compatible
XX CC
XX OPERATING SYSTEM: PC-DOS/MS-DOS
XX CC
XX SOFTWARE: PatentIn Release #1.0, Version #1.25
XX CC
XX CURRENT APPLICATION DATA:
XX CC
XX APPLICATION NUMBER: US/07/901,703
XX CC
XX FILING DATE: 19920616
XX CC
XX CLASSIFICATION: 435
XX CC
XX ATTORNEY/AGENT INFORMATION:
XX CC
XX NAME: PITCHER ESQ, EDMUND R
XX CC
XX REGISTRATION NUMBER: 27,829
XX CC
XX REFERENCE/DOCKET NUMBER: STK-057
XX CC
XX TELECOMMUNICATION INFORMATION:
XX CC
XX TELEPHONE: 617/248-7000
XX CC
XX INFORMATION FOR SEQ ID NO: 11:
XX CC
XX SEQUENCE CHARACTERISTICS:
XX CC
XX LENGTH: 402 amino acids
XX CC
XX TYPE: AMINO ACID
XX CC
XX TOPOLOGY: 1linear
XX CC
XX MOLECULE TYPE: protein
XX CC
XX SEQUENCE 402 AA; 44698 MW; 800499 CN;
XX SQ
XX
XX Query Match 1.8%; Score 125; DB 1; Length 402;
XX Best Local Similarity 39.7%; Pred.No. 5.57e-01;
XX Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;
XX
XX Db 12 GLIALCAGGGGPGRRPP-GC-PQRRIG-AERRDVQREILAVLGLPGR-PRRAPPAS 67
XX Qy 194 GPPLQLGAATQA-RPPHASGPRRRIGCERAMNHVREAGVPLGLPAPGARRRGGASR 252
XX
XX Db 68 RLP 70
XX Qy 253 SLP 255
XX
XX RESULT 11
XX ID US-08-147-023-29 STANDARD; PRT; 402 AA.
XX AC xxxxxx
XX DT
XX DE
XX Sequence 29, Application US/08147023
XX CC
XX Patent No. 5468845
XX CC
XX GENERAL INFORMATION:
XX CC
XX APPLICANT: OPPERMANN, HERMANN
XX CC
XX APPLICANT: OZKAYNAK, ENGİN
XX CC
XX APPLICANT: KUBERASAMPATH, THANGAVEL

```

CC APPLICANT: RUEGER, DAVID C.  
CC APPLICANT: PANG, ROY H.L.  
CC TITLE OF INVENTION: OSTEOGENIC DEVICES  
CC NUMBER OF SEQUENCES: 33  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: TESTA, HORWITZ & THIBEAULT  
CC STREET: 53 STATE STREET  
CC CITY: BOSTON  
CC STATE: MASSACHUSETTS  
CC COUNTRY: U.S.A.  
CC ZIP: 02109  
CC  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/147,023  
CC FILING DATE: 21-FEB-1992  
CC CLASSIFICATION: 530  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 810,560  
CC FILING DATE: 20-DEC-1991  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 827,052  
CC FILING DATE: 28-JAN-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 660,162  
CC FILING DATE: 22-FEB-1991  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 621,988  
CC FILING DATE: 04-DEC-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 621,849  
CC FILING DATE: 04-DEC-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 616,374  
CC FILING DATE: 21-NOV-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 600,024  
CC FILING DATE: 18-OCT-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 599,543  
CC FILING DATE: 18-OCT-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 579,865  
CC FILING DATE: 07-SEP-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 569,920  
CC FILING DATE: 20-AUG-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 483,913  
CC FILING DATE: 22-FEB-1990  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 422,613  
CC FILING DATE: 17-OCT-1989  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 315,342  
CC FILING DATE: 23-FEB-1989  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 232,630  
CC FILING DATE: 15-AUG-1988  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 179,460  
CC FILING DATE: 08-APR-1988  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: PITCHER, EDMUND R.  
CC REGISTRATION NUMBER: 27,829  
CC REFERENCE/DOCKET NUMBER: CRP-001CP6  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 617/248-7000  
CC TELEFAX: 617/248-7100  
CC  
CC INFORMATION FOR SEQ ID NO: 29:

```

CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 402 amino acids
CC      TYPE: amino acid
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: protein
CC      SEQUENCE 402 AA; 44698 MW; 800499 CN;
SQ
Query Match          1.8%; Score 125; DB 1; Length 402;
Best Local Similarity 39.7%; Pred.No.5.5/e-01;
Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5;
Db      12 GLATCALGGGPGRLRPP-GC-PPRRLG-ARERDVOREILAVLGLPGR-PPRRAPPAS 67
QY      194 GPLXQLGAAIQA-RPPHAGSGPRRRRLGCEANMHSYREAGVPLGLPAPGARRRGSSASR 252
Db      68 RLP 70
QY      253 SLP 255
RESULT 12
ID      US-08-643-563A-21          STANDARD;          PRT;          402 AA.
XX      xxxxxx
XX
XX
XX
XX
XX      Sequence 21, Application US/08643563A
DE      Sequence 21, Application US/08643563A
CC      Patent No. 5707810
CC      GENERAL INFORMATION:
CC      APPLICANT: SMART, JOHN
CC      APPLICANT: OPPERMAN, HERMAN
CC      APPLICANT: OZKAYNAK, ENGIN
CC      APPLICANT: KUBERASAMPATH, THANGAVEL
CC      APPLICANT: RUEGER, DAVID C.
CC      APPLICANT: PANG, ROY H.L.
CC      APPLICANT: COHEN, CHARLES M.
CC      TITLE OF INVENTION: MORPHOGENIC PROTEIN SCREENING METHOD
CC      NUMBER OF SEQUENCES: 33
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
CC      STREET: 45 SOUTH STREET
CC      CITY: HOPKINTON
CC      STATE: MA
CC      COUNTRY: USA
CC      ZIP: 01748
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Floppy disk
CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: PatentIn Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: US/08/643,563A
CC      FILING DATE: 06-MAY-1996
CC      CLASSIFICATION: 435
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: TWOMEY ESQ., MICHAEL J.
CC      REGISTRATION NUMBER: 38,349
CC      REFERENCE/DOCKET NUMBER: CRP-058CN2
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (508) 435-6951
CC      TELEFAX: (508) 435-6951
CC      INFORMATION FOR SEQ ID NO: 21:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 402 amino acids
CC      TYPE: amino acid
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: protein
CC      SEQUENCE 402 AA; 44764 MW; 803627 CN;
SQ
Query Match          1.8%; Score 125; DB 1; Length 402;

```







05-Sep-1997  
S03175; S03175; S10890; D25372; E38355; A03295; A61294;  
S62891

REFERENCE  
#authors Lyons, K.M.; Stein, J.H.; Smithies, O.  
#journal Genetics (1988) 120:267-278  
#title Length polymorphisms in human proline-rich protein genes  
generated by intragenic unequal crossing over.  
#cross-references MIM:89121440  
#accession S03176  
##status translation not shown  
##molecule\_type DNA  
##residues 35-310 ##label LY1  
##cross-references EMBL:X07715  
##note large allele  
#accession S03175  
##status translation not shown  
##molecule\_type DNA  
##residues 35-36, 'E', 38-112, 155-310 ##label LY2  
##cross-references EMBL:X07704  
##note medium allele  
#accession S10890  
##status preliminary; translation not shown  
##molecule\_type DNA  
##residues 1-39, 60-112, 'I', 114-115, 'P', 117-121, 185-271, 'A', 273-310  
##label LY3  
##cross-references EMBL:X07882; MID:935647; PID:9296670  
#accession A92492  
#authors Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.  
#journal J. Biol. Chem. (1985) 260:11123-11130  
#title Differential RNA splicing and post-translational cleavages in  
the human salivary proline-rich protein gene system.  
#cross-references MIM:85289325  
#accession D25372  
##molecule\_type mRNA  
##residues 1-36, 'E', 38-112, 'T', 114-115, 'P', 117-121, 185-271, 'A',  
273-310 ##label MAE

REFERENCE  
#authors Kauffman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
#journal Biochemistry (1991) 30:3351-3356  
#title Basic proline-rich proteins from human parotid saliva:  
relationships of the covalent structures of ten proteins  
from a single individual.  
#cross-references MIM:91190884  
#accession E38355  
##molecule\_type protein  
##residues 241-254, 'KN', 257-310 ##label KAU

REFERENCE  
#authors Salton, E.; Isemura, S.; Sanada, K.  
#journal J. Biochem. (1983) 93:495-502  
#title Complete amino acid sequence of a basic proline-rich peptide,  
P-D, from human parotid saliva.  
#cross-references MIM:83186122  
#accession A03295  
##molecule\_type protein  
##residues 241-310 ##label SAI

REFERENCE  
#authors Shimomura, H.; Kanai, Y.; Sanada, K.  
#journal J. Biochem. (1983) 93:857-863  
#title Amino acid sequences of glycopeptides obtained from basic  
proline-rich glycoprotein of human parotid saliva.  
#accession A61294  
##molecule\_type protein  
##residues 54-57, 'E', 59-73, 'R', 82-101 ##label SHI

REFERENCE  
#authors Chailton, A.J.; Baxter, N.J.; Lilley, T.H.; Haslam, E.;  
McDonald, C.J.; Williamson, M.P.  
#journal FEBS Lett. (1996) 382:289-292  
#title Tannin interactions with a full-length human salivary  
proline-rich protein display a stronger affinity than with  
single proline-rich repeats.  
#accession S62891  
##molecule\_type protein



```
#authors      Parker, B.D.; Bankier, A.; Satchwell, S.; Barrell, B.;
              Farrel, P.J.
#submission   Submitted to the EMBL Data Library, August 1990
#description   Sequence and transcription of Raj1 Epstein-Barr virus DNA
               spanning the B95-8 deletion region.
#accession    S27923
#status       Preliminary
##molecule-type DNA
##residues    1-924 ##label PAR
##cross-references EMBL:M35547; NID:g330420; PID:g330421
SUMMARY       #length 924 #molecular-weight 94304 #checksum 8997

Query Match          1.9%; Score 132; DB 2; Length 924;
Best Local Similarity 28.2%; Pred.No.5,19e-03;
Matches 37; Conservative 33; Mismatches 52; Indels 9; Gaps 9;

Db 38 AAPRAGPEPTRLQATPRRRSGAADPADPVGHPA-P-RAPGPEPTRLQ-PATPRRSQA 95
| : : : | | : : : : : | : : : : | : : : : | : : : : | : : : : |
Oy 206 ARPRHAGSPRRRLCCERAMNHSTREAGVPLGLPAFAKRRGGASASLPLRPFRGAA 255
::: ||| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Oy 266 PEPEPTPVGGSMAMPGRTRGSDRG-FCVVSPARP-AEATSLLEGALSTRSHSHP-SVG 322
::: ||| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 152 RLQPATPRRSQ 162
| : : : | : : :
Oy 323 RQHNGAPRSTS 333

RESULT 11
ENTRY C29149 #type fragment
TITLE proline-rich protein - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change
20-Mar-1998
ACCESSIONS C29149
REFERENCE A92501
#authors Clements, S.; Mehanasho, H.; Carlson, D.M.
#journal J. Biol. Chem. (1985) 260:13471-13477
#title Novel multigene families encoding highly repetitive peptide
        sequences. Sequence analyses of rat and mouse proline-rich
        protein cDNAs
#cross-references MUID:86033799
#contents Clone pUMP40
#accession C29149
##molecule-type mRNA
##residues 1-227 ##label CLE
CLASSIFICATION ##cross-references GB:M1902; NID:g200544; PID:g200545
                #superfamily proline-rich protein
SUMMARY        #length 227 #checksum 1378

Query Match          1.8%; Score 125; DB 2; Length 227;
Best Local Similarity 27.3%; Pred.No.3,67e-02;
Matches 41; Conservative 34; Mismatches 67; Indels 8; Gaps 8;

Db 49 QQGPPPGGPQRPRPGPPPGGFQPR-PDGPDPGGGPQRPRPGPPPGGFQPR-PQ 106
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Oy 199 QLGAATQARPHNAS-GRRRLGCE-RAWNHSYDEAGVPLCLPARGARRGGSSASRLPL 256
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 107 GERPPGGQRRPRPGPPPGG-PQRRPRGGPPPGGQLRRPQGRPPPARPQRRPQGRP 165
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Oy 257 PKRPRGAPAEERKTPVGGSMAMPGRTRGPSDRGFCVVSPAR-PAEAATSLLEGALSGTR 315
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 166 PPAGQPAPPO-GPPPTGPORPTQGP-PP 193
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Oy 316 HSHPSVGRQHNGAPRSTSRPRMDTQCP 345

RESULT 12
ENTRY D34768 #type complete
TITLE ORF4 protein - Orf virus (strain NZ2)
ORGANISM #formal_name Orf virus
DATE 23-Aug-1991 #sequence_revision 23-Aug-1991 #text_change
```

```

ACCESSIONS      31-Oct-1997
REFERENCE        D34768
#authors        Fraser, K.M.; Hill, D.F.; Mercer, A.A.; Robinson, A.J.
#journal        Virology (1990) 176:379-389
#title          Sequence analysis of the inverted terminal repetition in the
                genome of the parapoxvirus, orf virus.
#cross-references MUID:90266454
#accession      D34768
#status         preliminary
##molecule_type DNA
##residues      1-264 ##label FRA
##cross-references GB:M30023; EMBL:M37623; NID:g332561; PID:g332566
SUMMARY          #length 264 #molecular-weight 25613 #checksum 1957

Query Match      1.8%; Score 129; DB 2; Length 264;
Best Local Similarity 28.2%; Pred.No.1,2ie-02;
Matches          33; Conservative 40; Mismatches 37; Indels 7; Gaps 5;

Db    150 RAPGAGATPPAPPAASAAASRRHGRPARSSSPAERPPAPARAALAAPRAASS-GS 208
Qy    223 RAMNYSIV-EGVPLGLPAPARRRGGSASRSLPLPRP--R-RGAAPPEERTPVWQGS 277
Db    209 RASGGPPRSAR-SAASGSRPASGAPAAAPASARSASGAGARGCAPAGW 264
Qy    278 WAHPRTGSPDRGCVCVSPARPDEATSLGALSSTRHSHPVSVGROHHAGPFSTSR 334

RESULT   13
ENTRY    S19560      #type complete
TITLE    proline-rich protein MP4 - mouse
ORGANISM Mus musculus #common_name house mouse
DATE     22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
                20-Mar-1998
ACCESSIONS S19560; S22570
REFERENCE  S19560
#authors   Roberts, S.G.E.; Layfield, R.; Bannister, A.J.; McDonald, C.J.
#journal   Eur. J. Biochem. (1991) 202:969-974
#title     Gene sequence of mouse B-type proline-rich protein MP4.
            Transcriptional start point and an upstream phylogenetic
            footprint with ets-like and rel/NFkB-like elements.
#cross-references MUID:92111548
#accession S19560
#status    preliminary
##molecule_type DNA
##residues 1-300 ##label ROB
##cross-references GB:X58438; NID:g53181; PID:g53182
REFERENCE  S22570
#authors   Roberts, S.G.E.; Layfield, R.; McDonald, C.J.
#journal   Nucleic Acids Res. (1991) 19:5205-5211
#title     The mouse proline-rich protein Mpe promoter binds
            1soptenaline-inducible paroviral nuclear proteins via a
            highly conserved NFkB/rel-like site.
#cross-references MUID:92020206
#accession S22570
#status    preliminary
##molecule_type DNA
##residues 1-14 ##label RO2
##cross-references EMBL:X61126
CLASSIFICATION #superfamily proline-rich protein
SUMMARY        #length 300 #molecular-weight 31129 #checksum 8443

Query Match      1.8%; Score 131; DB 2; Length 300;
Best Local Similarity 25.2%; Pred.No.6,89e-03;
Matches          34; Conservative 40; Mismatches 54; Indels 7; Gaps 7;

Db    131 PCNOGCGPPPGC-GPOQRP-QPNNGCGPPPPGGQQNPORCGNGOGGRP-OGCPNP-PPR 166
Qy    202 RATQARPPPHASGPERRIGCCERAMNHVSREGVPLGLPAPARRRGGSASASLSLPLPRR 261
Db    187 PCNOGCGPPPGCPOOARLPNOPGNOCGPPPOGC-PQAPRPGNOCGPPGCGPGPPTGMOQ 245
Qy    262 RGAAEPERTPVGGCSWAHPRTGSPDRGCVCVSPARPA-EAATSLGALSCTRHSHP 320

```



\*\*\*\*\*  
MPSrch\_PP protein - protein database search, using Smith-Waterman algorithm  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Run on: Fri Dec 18 18:12:41 1998; Maspar time 25.51 seconds  
Tabular output not generated.

Title: >US-08-951-733-14  
Description: (1-949) from US08951733.pep (1 of 2)  
Perfect Score: 7113  
Sequence: 1 HASGQRCVLTWTEALAPAT.....PVEDMAGTAYQMPAHGL 949

Scoring table: PAM 150  
Gap 11

Searched: 74019 seqs, 26840295 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot35  
1:swissprot

Statistics: Mean 54.986; Variance 96.850; scale 0.568

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	162	2.3	660	1	YH1L_EBV	4.85e-09
2	153	2.2	504	1	ATIN_HSVBP	1.47e-07
3	150	2.1	234	1	PREM_HUMAN	4.48e-07
4	142	2.0	276	1	PRPL_HUMAN	8.26e-06
5	133	1.9	247	1	PRP4_HUMAN	1.97e-04
6	137	1.9	628	1	V7OK_TYMC	4.88e-05
7	125	1.8	261	1	PRP2_MOUSE	2.95e-03
8	131	1.8	296	1	PRP3_MOUSE	3.91e-04
9	126	1.8	302	1	NOD3_RHILP	2.12e-03
10	125	1.8	402	1	BMP8_HUMAN	2.95e-03
11	125	1.8	501	1	MEH1_HUMAN	2.95e-03
12	127	1.8	566	1	TS13_MOUSE	1.51e-03
13	121	1.7	322	1	NOD1_RHISN	1.09e-02
14	118	1.7	399	1	BMDA_MOUSE	2.87e-02
15	118	1.7	431	1	UL61_HCMVA	2.87e-02
16	118	1.7	514	1	VE2_HPV05	2.87e-02
17	114	1.6	202	1	VG01_HSVBP	1.01e-01
18	116	1.6	270	1	Y091_NPVOP	5.39e-02
19	115	1.6	279	1	Y091_NPVOP	7.38e-02
20	114	1.6	321	1	NOD1_BRAJA	1.01e-01
21	115	1.6	331	1	PRP1_HUMAN	7.38e-02
22	114	1.6	408	1	CGE1_XENLA	1.01e-01
23	114	1.6	408	1	CGE2_XENLA	1.01e-01

24	111	1.6	498	1	VE2_HPV08	2.52e-01
25	111	1.6	509	1	VE2_HPV36	2.52e-01
26	113	1.6	514	1	VE2_HPV58	1.37e-01
27	112	1.6	633	1	LA17_YEAST	1.86e-01
28	113	1.6	704	1	SYN1_RAT	1.37e-01
29	114	1.6	705	1	SYN1_HUMAN	1.01e-01
30	113	1.6	1255	1	MGCL_HUMAN	1.37e-01
31	115	1.6	3149	1	TEGU_EBV	7.38e-02
32	107	1.5	174	1	BAR1_CHITE	8.32e-01
33	106	1.5	233	1	YFOR_ECOLI	1.11e+00
34	106	1.5	285	1	YAFY_ECOLI	1.11e+00
35	109	1.5	301	1	NOD3_RHILP	4.60e-01
36	107	1.5	318	1	NOD3_RHILP	8.32e-01
37	109	1.5	324	1	MAPD_NEUCR	4.60e-01
38	110	1.5	408	1	CGE3_XENLA	3.41e-01
39	110	1.5	494	1	MEH1_MOUSE	8.32e-01
40	107	1.5	529	1	DNB2_ADE05	8.32e-01
41	110	1.5	628	1	V7OK_TYMC	3.41e-01
42	108	1.5	706	1	SYN1_BOVIN	6.20e-01
43	105	1.5	863	1	HTS2_NEUCR	1.49e+00
44	105	1.5	890	1	ACOL1_ECOLI	1.49e+00
45	105	1.5	2774	1	MAPA_RAT	1.49e+00

## ALIGNMENTS

RESULT	ID	Query Match	Standard	PRT	Length
1	YH1L_EBV	48; Conservative	34; Mismatches	60; Indels	12; Gaps
AC	P03181	2.3%; Score 162; DB 1; Length 660;			
DT	21-JUL-1986 (REL. 01, CREATED)				
DT	21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)				
DT	15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)				
DE	HYPOTHETICAL BHLP1 PROTEIN.				
OS	EPSTEIN-BARR VIRUS (STRAIN B95-8) (HUMAN HERPESVIRUS 4).				
OC	VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; GAMMAHERPESVIRINAE.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 84270667.				
RA	BAER R., BANKIER A.T., BIGGIN M.D., DEININGER P.L., FARRELL P.J.,				
RA	GLIBSON T.J., HATFIELD G., HUDSON G.S., SATCHWELL S.C., SEGUN C.,				
RL	TUFFENL P.S., BARRELL B.G.,				
RL	NATURE 310:207-211(1984).				
DR	EMBL: V01555; -. NOT_ANNOTATED_CDS.				
DR	PIR: A03742; Q0283.				
KW	HYPOTHETICAL PROTEIN; EARLY PROTEIN; REPEAT.				
FT	DOMAIN 149 648 4 X 125 AA TANDEM REPEATS.				
FT	REPEAT 149 273 1.				
FT	REPEAT 274 398 2.				
FT	REPEAT 399 523 3.				
FT	REPEAT 524 648 4.				
SO	SEQUENCE 660 AA; 66244 MW; 372F08C5 CRC32;				
Db	251 GPPTRSGAAQ-RTHRRPCCPSARNPCCPTWR--R-SGAQRGHPGAGQDSGP 306				
Qy	194 GPPYQLQAATQAPPPHNSG-PRR-R-LGGERAMNHSVRAGVPLGIPAGG-RRGGS 249				
Db	307 TGGPAPAGATGTAACGPGGAAPVSGATPHRRSGGAPDP--AAARLPPEQDRLP 364				
Qy	250 ASRLSPLEKRRRGAADPER-TVGGGSMHNPGRTPSDRGCVVSPARPAEATSL 308				
Db	365 QDLAAQRC-PAGPPTRSGA-AAQTRH-R-RPCCP 397				
Qy	309 GALSGRSHSVGRHAGGPPSTSRPPMDTDCP 344				
RESULT	2	STANDARD;	PRT;	504 AA.	
ID	ATIN_HSVBP				
AC	P30020;				
DT	01-APR-1993 (REL. 25, CREATED)				

DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)  
 DE ALPHA TRANS-INDUCING PROTEIN (ALPHA-TIP).  
 OS BOVINE HERPESVIRUS TYPE 1 (STRAIN P8-2).  
 OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; ALPHAHERPESVIRINAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 93012995.  
 RA CARPENTER D.E., MISRA V.;  
 RL GENE 119:259-263(1992).  
 CC -1- FUNCTION: RESPONSIBLE FOR TRANSCRIPTIONAL ACTIVATION OF IMMEDIATE  
 CC -1- PROMOTERS (ALPHA GENES) (BY SIMILARITY).  
 CC -1- SIMILARITY: TO OTHER HERPESVIRUSES ALPHA TRANS-INDUCING PROTEIN.  
 DR EMBL: Z11610.E264419; .  
 DR PIR: S24329; S24329.  
 DR PIR: JCI306; JCI306.  
 KW TRANSCRIPTION REGULATION; TRANS-ACTING FACTOR; DNA-BINDING.  
 QO SEQUENCE 504 AA; 54028 MW; 79F42020 CRC32;

Query Match	2.28;	Score 153;	DB 1;	Length 504;
Best Local Similarity	31.38;	Pred. No. 1.47e-07;		
Matches	46;	Conservative 33;	Mismatches 56;	Indels 12;
				Gaps 11;

[illegible]

RESULT	3	STANDARD;	PRT;	234	AA.
ID	PREM_HUMAN				
AC	P10161; P02813;				
DT	01-MAR-1989 (REL. 10, CREATED)				
DT	01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)				
DT	01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)				
DE	SALIVARY PROLINE-RICH PROTEIN PO (ALLELE M) (CONTAINS: PEPTIDE P-D) (FRAGMENT).				

OC	HOMO SAPIENS (HUMAN).
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA;
OC	EUTHERIA; PRIMATES.
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE: 89121439.
RA	LYONS K.M., STEIN J.H., SMITHIES O.;
RL	GENETICS 120:235-265(1988).
RN	[2]
RP	SEQUENCE OF 165-234.
RX	MEDLINE: 83186122.
RA	SATCHO E., ISEMURA S., SANADA K.;
RL	J. BIOCHEM. 93:495-502(1983).
DR	EMBL: X07704; E265547; -.
DR	PIR: A03295; PIHUSD.
DR	PIR: S03175; S03175.
DR	HSSP: P19999; ICLG.
DR	MIM: 168730; -.
DR	MIM: 180990; -.
KW	REPEAT; PAROTID GLAND; SALIVA; MULTIGENE FAMILY.
FT	NON_TER
FT	1
FT	CHAIN
FT	165
FT	234
FT	234 AA; 23676 MW; ED2D4ADC CRC32;
FT	PEPTIDE P-D.

Query Match	2.18;	Score 150;	DB 1;	Length 234;
Best Local Similarity	29.38;	Pred. No. 4.48e-07;		
Matches	43;	Conservative	41;	Mismatches 53;
			Indels 10;	Gaps 10;

[illegible]

RESULT	4	STANDARD;	PRT;	276 AA.
ID	PRPL.HUMAN			
AC	P10162; P02813;			
DT	01-MAR-1989 (REL. 10, CREATED)			
DT	01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)			
DT	01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)			
DE	SALIVARY PROLINE-RICH PROTEIN P0 (ALLELE K) (CONTAINS: PEPTIDE P-D)			
DE	(PRAGMENT) .			
GN	PRB4.			
OS	HOMO SAPIENS (HUMAN).			
OC	EUDAROTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;			
OC	EUTHERIA; PRIMATES.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 89121439.			
RA	LYONS K.M., STEIN J.H., SMITHIES O.;			
RL	GENETICS 120:235-265(1988).			

RA	SEQUENCE OF 207-276.		
RP			
RQ	MEDLINE; 8318612.		
SA	SATO H E., ISEKURA S., SANADA K.;		
SL	J. BIOCHEM. 93:495-502(1983).		
RL	EMBL; X07715; E4806; ALT_SEQ.		
DR	PIR; A03295; PIHSD.		
DR	PIR; S03176; S03176.		
DR	HSSP; P19999; 1CLG.		
DR	MTM; 166730; -.		
DR	MTM; 180990; -.		
KW	REPEAT; PAROTID GLAND; SALIVA;		
FT	NON_TER	1	
FT	CHARIN	207	276
FT		1	
QO	SEQUENCE	276 AA;	27616 MW; 4838945A CRC32;
			PEPTIDE P-D.

Query Match	2.0%;	Score 142;	DB 1;	Length 276;
Best Local Similarity	28.28;	Pred. No. 8.26e-06;		
Matches	40;	Conservative	41;	Mismatches 52;
			Indels	9;
			Gaps	9

[illegible]

ID	PREP_HUMAN	STANDARD:	PRT:	247 AA.
AC	P10163; P02813;			
DT	21-JUL-1986 (REL. 01, CREATED)			
DT	13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)			
DT	01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)			
DE	SALIVARY PROLINE-RICH PROTEIN PO PRECURSOR (ALLELE S).			
GN	PRB4.			
OC	HOMO SAPIENS (HUMAN).			
OC	EURARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA.			

Db 157 GVLLETETKPTFVSQVROKATGPRSPRPILLKRY-VH-VHDDPPSSILSRSSROLOPT 21  
Qy 194 GPPLYGQGAAGARPPHAA-GPR-RRGCGRRANNHSTRKNGVPLGLAPAPAR-RR-GGS 245  
Db 215 VRRPLAPNQHSHSPQPPPLDDDEGILGPRLARLHSTDDPPRR1-TPGPN-THDLRPL 272  
Qy 250 ASRSLPLPKRRPRGCAAPDEPRPVG-QGSA-HGQRTGSPSDRGFCVWSPAPRPEEATSL 307

DT 13-AUG-1987 (REL. 05, CREATED)  
DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)  
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)  
DE PROLINE-RICH PROTEIN MP-3 (FRAGMENT).  
GN PRP.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; TETRAPODA; MAMMALIA,  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 86059475.  
RA ANN D.K., CARLSON D.M.;  
RL J. BIOL. CHEM. 260:15863-15872(1985).  
DR EMBL; M12100; G200549; -.  
DR HSSP; P19999; 1CLG.  
KW REPEAT; SALIVA.  
FT NON\_TER 1

[illegible]

```

DE BONE MORPHOGENETIC PROTEIN 8 PRECURSOR (BMP-8) (OSTEOGENIC PROTEIN 2)
DE (OP-2).
GN BMP8.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-HIPOCAMPUS.
RX MEDLINE; 93094231.
RA OEKANAK E., SCHNEGELSBERG P.N.J., JIN D.F., CLIFFORD G.M.,
RA WARREN F.D., DRIER E.A., OPPERMANN H.;
RL J. BIOL. CHEM. 267:25220-25227(1992).
CC -1- FUNCTION: INDUCES CARTILAGE AND BONE FORMATION. MAY BE THE
CC OSTEOINDUCTIVE FACTOR RESPONSIBLE FOR THE PHENOMENON OF
CC EPITHELIAL HOMEOSTASIS (BY SIMILARITY).
CC AND BONE HOMEOSTASIS (BY SIMILARITY).
CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED.
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
DR EMBL; M97016; G189390; -.
DR PIR; A45056; A45056.
DR HSSP; P08112; ITFG.
DR MIM; 602284; -.
DR PROSITE; PS00250; TGF_BETA; 1.
KW SIGNAL. GROWTH FACTOR: CYTOKINE. BONE: CARTILAGE; GLYCOPROTEIN.
FT SIGNAL. 1 19 POTENTIAL.
FT PROPEP 20 263 POTENTIAL.
FT CHAIN 264 402 BONE MORPHOGENETIC PROTEIN 8.
FT DISULFID 301 367 BY SIMILARITY.
FT DISULFID 330 399 BY SIMILARITY.
FT DISULFID 334 401 BY SIMILARITY.
FT DISULFID 366 366 INTERCHAIN (BY SIMILARITY).
FT CARBOHYD 158 158 POTENTIAL.
FT CARBOHYD 343 343 POTENTIAL.
SQ SEQUENCE 402 AA; 44764 MW; 8F20C81B CRC32;

Query March 1.8% Score 125; DB 1; Length 402;
Best Local Similarity 39.7% Pred. No. 2,95e-03;
Matches 25; Conservative 12; Mismatches 21; Indels 5; Gaps 5

Db 68 RLP 70
OY 253 SLP 255

RESULT 11
ID MFH1_HUMAN STANDARD; PRT; 501 AA.
AC Q99958;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE MESENCHYME FORK HEAD PROTEIN 1 (MFH-1 PROTEIN) (TRANSCRIPTION FACTOR
DE FKHL14).
DE FKHL14 OR MFH1.
GN HOMO SAPIENS (HUMAN).
OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97312712.
RA MIURA N., IIDA K., KAKIMURA H., YANG X.-L., SUGIYAMA T.;
RL GENOMICS 41:489-492(1997).
CC -1- FUNCTION: MIGHT BE INVOLVED IN THE FORMATION OF SPECIAL
CC MESENCHYMAL TISSUES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
DR EMBL; Y08823; E303016; -.
DR MIM; 602402; -.
DR PROSITE; PS00657; FORK_HEAD_1; 1.

```

[illegible]

RESULT	ID	MODL	PHISN	STANDARD;	PRT;	322 AA.
AC	P53559:	01-NOV-1997	(REL. 35, CREATED)			
DT	01-NOV-1997	(REL. 35, LAST SEQUENCE UPDATE)				
DT	01-NOV-1997	(REL. 35, LAST ANNOTATION UPDATE)				
DE	MODULATION PROTEIN D I.					
GN	NODDI OR Y4AL.					
OS	RHIZOBIUM SP. (STRAIN NGR234).					
OG	PLASMID STM PNG8234A.					
OC	PROKARYOTA: GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;					
OC	RHIZOBIACEAE.					
RA	[1]					
RA	SEQUENCE FROM N.A.					
RK	MEDLINE; 97305956.					
RL	FRIBERG C.A., FELLAY R., BAIRROCH A., BROUGHTON W.J., ROSENTHAL A.,					
RL	PERRET X.;					
RL	NATURE 387:394-401(1997).					
CC	-1- FUNCTION: NODD REGULATES THE EXPRESSION OF THE NOD ABCPE GENES					
CC	WHICH ENCODE OTHER NODULATION PROTEINS. NODD IS ALSO A NEGATIVE					
CC	REGULATOR OF ITS OWN EXPRESSION. BINDS FLAVONOIDS AS INDUCERS.					
CC	-1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL					
CC	REGULATORS.					
DR	EMBL; AE000065; G21823301; -					
DR	PROSITE; PS00044; HTH_LYSR_FAMILY; 1.					
KW	NODULATION; TRANSCRIPTION REGULATION; DNA-BINDING; ACTIVATOR;					
KW	REPRESSOR; MULTIGENE FAMILY; PLASMID.					
FT	DNA_BIND 23 42 H-T-H MOTIF (BY SIMILARITY).					
FT	SEQUENCE 322 AA; 36460 MW; FA5238EE CRC32;					
SO						
	Query Match	1.7%	Score 121;	DB 1;	Length 322;	
	Best Local	Similarity 30.1%;	Pred. No.1.09e-02;			
	Matches	28; Conservative	23; Mismatches 35;	Indels 7;	Gaps	
Db	44 LTRYDEL-FTMGRELIVPTP-R-AEALAPAVEALHILHLSITSDPNEAOSDRSFR 100					
Oy	36 LRSRREVLPATIVTRLRGPGWMLYQNGDPNARALVAOGLVCPMDA-RPPAARSFR 94					
Db	101 -II-LSDFMTLFLERVVYVAREAPAVSEFL 131					
Oy	95 QVSLCKELVARY-LORLCERGAKNVLAFGFALL 126					
RESULT	14					
ID	BMA8_MOUSE	STANDARD;	PRT;	399 AA.		
AC	P34821:					
DT	01-FEB-1994 (REL. 28, CREATED)					
DT	01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)					
DT	01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)					
DE	BONE MORPHOGENETIC PROTEIN 8A PRECURSOR (BMP-8A) (OSTEOGENIC PROTEIN					
DE	2) (OP-2).					
GN	BMP8A OR BMP8 OR BMP-8.					
OS	MUS MUSCULUS (MOUSE).					
OC	EUMAROTA: METAZOA: CHORDATA: VERTEBRATA; TETRAPODA; MAMMALIA;					
OC	EUTHERIA: RODENTIA.					
OC	[1]					
RN	SEQUENCE FROM N.A.					
RP	TISSUE-EMBRYO;					
RX	MEDLINE; 93094231.					
RA	OEKAKINAK E., SCHNEGBELSBURG P.N.J., JIN D.F., CLIFFORD G.M.,					
RA	WARREN F.D., DRIER E.A., OPPERMAN H.;					
RL	J. BIOL. CHEM. 267:25220-25227(1992).					
CC	-1- FUNCTION: INDUCES CARTILAGE AND BONE FORMATION. MAY BE THE					
CC	OSTEOINDUCTIVE FACTOR RESPONSIBLE FOR THE PHENOMENON OF					
CC	EPITHELIAL OSTEOGENESIS. PLAYS A ROLE IN CALCIUM REGULATION					
CC	AND BONE HOMEOSTASIS (BY SIMILARITY).					
CC	-1- SUBUNIT: HOMODIMER. DISULFIDE-LINKED.					
CC	-1- DEVELOPMENTAL STAGE: EXTENSIVE EXPRESSION FOUND IN 8-DAY EMBRYOS.					
CC	-1- FELLD DRASITICALLY IN 10-DAY EMBRYOS AND VIRTUALLY ABSENT IN 17-					
CC	DAY EMBRYOS.					
CC	-1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.					

DR EMBL: M97017; G200141; -  
 DR MGD: MGI.104515; BMP8A.  
 DR PROSITE: PS00250; TGF\_BETA\_1.  
 KW SIGNAL; GROWTH\_FACTOR; CYTOKINE; BONE; CARTILAGE; GLYCOPROTEIN.  
 FT SIGNAL 1  
 FT PROPEP 20 260  
 FT CHAIN 261 399  
 FT DISULFID 298 364  
 FT DISULFID 327 396  
 FT DISULFID 331 398  
 FT DISULFID 363 363  
 FT CARBOHYD 155 155  
 FT CARBOHYD 340 340  
 SO SEQUENCE 399 AA; 44764 MW; 7075A1ED CRC32;

Query Match 1.7%; Score 118; DB 1; Length 399;  
 Best Local Similarity 38.7%; Pred. No. 2.87e-02;  
 Matches 24; Conservative 12; Mismatches 22; Indels 4; Gaps 4;

Db 12 GLACALGGGHGPRP-HTC-PORRIG-ARERRMOREILLAVLGLPGR-PPRPAQPAAR 67  
 QY 194 GPPLYQLGAATQAPRPHASGPRRRRLGCERAMHNSVREAGVPLGLPAPGARRRGSASRS 253

Db 68 QP 69  
 QY 254 LP 255

RESULT 15  
 ID UL61\_HCMVA STANDARD; PRT; 431 AA.  
 AC P16818;  
 DT 01-AUG-1990 (REL. 15, CREATED)  
 DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL PROTEIN UL61.  
 GN UL61.  
 OS HUMAN CYTOMEGALOVIRUS (STRAIN AD169).  
 OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; BETAHERPESVIRINAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 90269039.  
 RA CHEE M.S., BANKIER A.T., BECK S., BOHNT R., BROWN C.M., CERNY R.,  
 RA HORSNELL T., HUTCHISON C.A. III, KOZARIDES T., MARTIGNETTI J.A.,  
 RA PREDIE E., SATCHWELL S.C., TOMLINSON P., WESTON K.M., BARRELL B.G.,  
 RL CURR. TOP. MICROBIOL. IMMUNOL. 154:125-169(1990).  
 DR EMBL: X17403; E27280; -  
 DR FJR; S09824; S09824.  
 KW HYPOTHETICAL PROTEIN.  
 SO SEQUENCE 431 AA; 44309 MW; 232AB9D7 CRC32;

Query Match 1.7%; Score 118; DB 1; Length 431;  
 Best Local Similarity 32.1%; Pred. No. 2.87e-02;  
 Matches 27; Conservative 20; Mismatches 31; Indels 6; Gaps 6;

Db 268 RGGGKPPPLGSPRA-TDGNRDPG-A-GVPAPGRMGSSGSGRGTPGRGPERRAAPGARP 324  
 QY 211 HASGPRRRRLGCERAMHNSVREAGVPLGLPAPGARRRGSAS-RSLPLKPRRRGAPEPE 269  
 Db 325 TAPDAPGRWDGPAADGAPAGLGRG 348  
 QY 270 RTPVG-OGSWAHFGRTRGPS-DRG 291

Search completed: Fri Dec 18 18:14:00 1998  
 Job time : 79 secs.



QY 880 RRDGLLRLVDDFLVTPHLTHAKTFLRTLVKGVPEYGCYVNLKRTVVPVEDEALGCT 939

Db 918 A 918

QY 940 A 940

RESULT 2

ID 014783 PRELIMINARY; PRT: 1132 AA.

AC 014783;

DT 01-JAN-1998 (TREMBLREL. 05, CREATED)

DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)

DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)

DE TELOMERASE CATALYTIC SUBUNIT.

GN HEST2.

OS HOMO SAPIENS (HUMAN).

OC EURAROTIA; METAACOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; PRIMATES.

RN [1]

RA SEQUENCE FROM N.A.

RA MEDLINE; 97433088.

RA MEYERSON M., COONTER C.M., EATON E.N., ELLISEN L.M., STEINER P.,

RA CADDE S.D., ZIAGRA L., BEIJERSENGEN R.L., DAVIDOFF M.J., LIU Q.,

RA BACCETTI S., HABER D.A., WEINBERG R.A.;

RL CELL. 90:785-795(1997).

DR EMBL; AF018167; G2347129; -

SO SEQUENCE 1132 AA; 126937 MW; C1E5E2AF CRC32;

Query Match 100.0%; Score 2214; DB 4; Length 1132;

Best Local Similarity 100.0%; Pred. No. 0.00e+00;

Matches 301; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 618 TSLRLFIKPGGLRPIVNDYVVGARTFRREKRAERLTSRYKALFSVYNTERRARPPGLG 677

QY 640 TSLRLFIKPGGLRPIVNDYVVGARTFRREKRAERLTSRYKALFSVYNTERRARPPGLG 699

Db 678 ASVGLDDIHRAMRTFVLRVADQPPPELYFYKVDTGAYDIPDDRLEVIASIIKPPON 737

QY 700 ASVGLDDIHRAMRTFVLRVADQPPPELYFYKVDTGAYDIPDDRLEVIASIIKPPON 759

Db 738 TYCVRRYAVVOKAAGHVRKAFKSHVSTLTDLPYMRQFVAHLQETSPLRDAVIEQSSS 797

QY 760 TYCVRRYAVVOKAAGHVRKAFKSHVSTLTDLPYMRQFVAHLQETSPLRDAVIEQSSS 819

Db 798 LNEASSGFLDFPLRFMCHAVIRKRSYVOCGIPGSIITSLCLCYGDMENKLFAGI 857

QY 820 LNEASSGFLDFPLRFMCHAVIRKRSYVOCGIPGSIITSLCLCYGDMENKLFAGI 879

Db 858 RRDGLLRLVDDFLVTPHLTHAKTFLRTLVKGVPEYGCYVNLKRTVVPVEDEALGCT 917

QY 880 RRDGLLRLVDDFLVTPHLTHAKTFLRTLVKGVPEYGCYVNLKRTVVPVEDEALGCT 939

Db 918 A 918

QY 940 A 940

RESULT 3

ID 070372 PRELIMINARY; PRT: 1122 AA.

AC 070372;

DT 01-AUG-1998 (TREMBLREL. 07, CREATED)

DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)

DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)

DE TELOMERASE REVERSE TRANSCRIPTASE.

OS MUS MUSCULUS (MOUSE).

OC EURAROTIA; METAACOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; RODENTIA.

RN [1]

RA SEQUENCE FROM N.A.

RA GREENBERG R.A., ALLSOPP R.C., CHIN L., MORIN G.B., DEPINHO R.A.;

RA ONCOGENE 0:0-0(1998).

DR EMBL; AF051911; G3005592; -

KN RNA-DIRECTED DNA POLYMERASE.

SO SEQUENCE 1122 AA; 127977 MW; 222075D6 CRC32;

Query Match 72.1%; Score 1596; DB 11; Length 1122;

Best Local Similarity 65.6%; Pred. No. 0.00e+00;

Matches 198; Conservative 68; Mismatches 33; Indels 3; Gaps 2;

Db 610 RLRFIPKPNGLRPIVNVSYMGSTFALGRKROAOHFQRLTKTSMLEVERKHPHMLGSS 669

QY 642 RLRFIPKPNGLRPIVNVSYMGSTFALGRKROAOHFQRLTKTSMLEVERKHPHMLGSS 701

Db 670 VLGNADYRTWRFAVLVRLADQTPRMYFYKADYTGAYDAIPGKLEVEVYANMIRHSEST 739

QY 702 VLGNADYRTWRFAVLVRLADQTPRMYFYKADYTGAYDAIPGKLEVEVYANMIRHSEST 760

Db 730 YCIRQIYAVVRDQGVHKSFRROYTTLSDLPYMGOLFHLQDSDASALNSVYIQST 789

QY 761 YCIRQIYAVVRDQGVHKSFRROYTTLSDLPYMGOLFHLQDSDASALNSVYIQST 818

Db 790 SMNSSSLPDPFLFRLHSHVYIGDRCTYOCOGIPGSIITSLCLCYGDMENKLFAG 849

QY 819 SMNSSSLPDPFLFRLHSHVYIGDRCTYOCOGIPGSIITSLCLCYGDMENKLFAG 878

Db 850 VORDGLLRLVDDFLVTPHLTHAKTFLRTLVKGVPEYGCYVNLKRTVVPVEDEALGCT 909

QY 879 VORDGLLRLVDDFLVTPHLTHAKTFLRTLVKGVPEYGCYVNLKRTVVPVEDEALGCT 938

Db 910 A 911

QY 939 TA 940

RESULT 4

ID 013339 PRELIMINARY; PRT: 988 AA.

AC 013339;

DT 01-JAN-1998 (TREMBLREL. 05, CREATED)

DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)

DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)

DE TELOMERASE REVERSE TRANSCRIPTASE 1.

GN SCHIZOSACCHAROMYCES POMBE (FUSION YEAST).

OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.

RN [1]

RA SEQUENCE FROM N.A.

RA STRAIN-972H-1.

RA MEDLINE; 97400623.

RA NAKAMURA T.M., MORIN G.B., CHAPMAN K.B., WEINRICH S.L., ANDREWS W.H.,

RA LINGNER J., HARLEY C.B., CECCH T.R.;

RL SCIENCE 277:955-959(1997).

RN [2]

RA SEQUENCE FROM N.A.

RA STRAIN-972H-1.

RA LINE M., RAJANDREAM M.A., BARRELL B.G., VOLCKAERT G.;

RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; AF015783; G2340168; -

DR EMBL; AL022299; E1285360; -

DR PRAM; PR00078; YVT

KN RNA-DIRECTED DNA POLYMERASE.

SO SEQUENCE 988 AA; 116328 MW; 39C385A7 CRC32;

Query Match 16.9%; Score 375; DB 3; Length 988;

Best Local Similarity 30.7%; Pred. No. 3.34e-52;

Matches 65; Conservative 51; Mismatches 90; Indels 6; Gaps 5;

Db 585 YFVADIRKSCYDRIKQDLMFRIYKRLKDE-FVIRKYATIH-ATSDRAATNFVSEAFSTY 642

QY 729 YFVADIRKSCYDRIKQDLMFRIYKRLKDE-FVIRKYATIH-ATSDRAATNFVSEAFSTY 788

Db 643 FDMYPERK--VQLLSKRT--SDILFVDVFDYWKSSSEIFKMLKEHLSGHIVKGNQYL 699

QY 789 TDLOPYMROFAHQLQETSPLRDAVIEQSSSLNEASSGFLDFPLRFMCHAVIRKRSYV 848

Db 700 QKVGIPGSIITSLCLCYGDMENKLFAGI 759

```

QY 849 QCGGIPGSSILSTLCSLCYGMENKLFAGIRRDG-LLRLVDEFLVTPHLTHAKTELR 907
DB 760 LSLRFEKHNFSTLEKTVINPENGNIINT 791
QY 908 TLVRGVPETGCVNLRKTVNPFVEDEALGGT 939

RESULT 5
ID 013338 PRELIMINARY; PRT: 989 AA.
AC 013338.
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE TELOMERASE REVERSE TRANSCRIPTASE 1.
GN TRT1
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972H-.
RX MEDLINE; 97400623.
RA NAKAMURA T.M., MORIN G.B., CHAPMAN K.B., WEINRICH S.L., ANDREWS W.H.,
RA LINGNER J., HARLEY C.B., CECCH T.R.;
RL SCIENCE 277:955-959(1997).
DR EMBL: AF015783; G2340169; -.
RN PFAM: PF00078; rvt.
KM RNA-DIRECTED DNA POLYMERASE.
SQ SEQUENCE 989 AA; 116456 MW; FDE74202 CRC32;

Query Match 16.9%; Score 375; DB 3; Length 989;
Best Local Similarity 30.7%; Pred. No. 3,34e-52;
Matches 65; Conservative 51; Mismatches 90; Indels 6; Gaps 5;

DB 586 YFVRIDKSCYDRKIDMERIVKKLKDP-EVIRKATIH-ATSDRATKNFSEAFSY 643
QY 729 YFVYVDYTGAVDTIPDRLTETIASIIRPONTYCVRAVAVVQKAHGHVRAFKSHVSTL 788
DB 644 FDMVPEEK--VVOLISMT-SDLEFVVDVDTWTSSEIFMKLKHLSGHVYKIGNSYL 700
QY 789 TDLQYMQVYAHLOETSPRLDAVYIEOSSSLNENSSGLFVFLRMCHEHVRIRKGSYV 848
DB 701 QKVGIPGSSILSTLCSLCYGMENKLFAGIRRDG-LLRLVDEFLVTPHLTHAKTELR 760
QY 849 QCGGIPGSSILSTLCSLCYGMENKLFAGIRRDG-LLRLVDEFLVTPHLTHAKTELR 907
DB 761 LSLRFEKHNFSTLEKTVINPENGNIINT 792
QY 908 TLVRGVPETGCVNLRKTVNPFVEDEALGGT 939

RESULT 6
ID 000939 PRELIMINARY; PRT: 1031 AA.
AC 000939.
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE TELOMERASE SUBUNIT P13.
OS EUKARYOTA; MITOCHONDRIAL EUKARYOTES; ALVEOLATA; CILIOPHORA;
OC EUKARYOTA; MITOCHONDRIAL EUKARYOTES; ALVEOLATA; CILIOPHORA;
OC HYPOTRICH; EUPLOTIDA; EUPLOTES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97274210.
RA LINGNER J., HUGHES T.R., SHEVCHENKO A., MANN M., LONDEBLAD V.,
RA CECCH T.R.;
RL SCIENCE 276:561-567(1997).
DR EMBL: U95964; G2072336; -.
SQ SEQUENCE 1031 AA; 122562 MW; 21A885CD CRC32;

Query Match 9.1%; Score 201; DB 5; Length 1031;
Best Local Similarity 24.3%; Pred. No. 1,81e-17;
Matches 77; Conservative 82; Mismatches 126; Indels 32; Gaps 21;

```

```

DB 512 GKLRIPKKTFRP--MTF--NKKIVSDRKTTLTNTKLNSHMLTKLRMFKDP 567
QY 641 SRLFIKPGDLRIYVMDYVVGARTERRERKARLTSRKALFS--VLNERARR-PGL 697
DB 568 FGAVFNDVDMKRYEEFVCK-WKQVGPKLFEATMDIEKQYSDVNRKLSFTLTKLL 626
QY 698 LGASVLGLDDIHRAMRTFVLVRAQDDPELXFVKYDVTGAVDTIPDRLETVI-AS-II 755
DB 627 SSDFWMTAQILKRNKNIVYDSKMFRRKEMKDYFRQKALLEGQYPIFLSVLENDQ 686
QY 756 KPQ--NYCY--RRYAVV-Q-KAAGHVRA-KKSHVSTL-DLQPYMRF-V-AHLQ 804
DB 687 DLNKKTLIVAKRNFYKONLQPIYIN-ICOYNTINENKFEYKQKGIPOGICVSSIL 745
QY 805 TSPRLDAVYIEOSSSLNENSSGLFVFLRMCHEHVRIRKGSYVQCGGIPGSSILSTL 863
DB 746 SSFYATLLESSLGFLDESDNPNPNVNLRLTLDYLLTTOENNAVLEIEKLINVS 805
QY 864 CSLCYGDMENK-L-F--AGIRRDG---LLRLVDEFLVTPHLTHAKTELRVLVGV 914
DB 806 ENGRFNNKKLQTSFPL 822
QY 915 FYGCVNLRKTVNPFV 931

```

```

RESULT 7
ID 006163 PRELIMINARY; PRT: 884 AA.
AC 006163.
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE CHROMOSOME XII COSMID 8543.
GN LB543.12.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C (AB972);
RX MEDLINE; 97313267.
RA JOHNSTON M., HILLIER L., RILES L., ALBERMANN K., ANDRE B., ANSORGE W.,
RA BERNES V., BRUCKNER M., DELIUS H., DUBOIS E., DUSTERHOFT A.,
RA ENTIAN K.D., FLOETH M., GOFFEAU A., HEBLING U., HEUMANN K.,
RA HEISS-NEITZEL D., HILBERT H., HILGER F., KLEINE K., KOTTER P.,
RA LOUIS E.J., MESSENGY F., MEMES H.W., MIOSGA T., MOSTL D.,
RA MULLER-AUER S., NENTWICH U., OBERMAYER B., PIVAVANDI E., POHL T.M.,
RA PORTERLE D., PURNELLE B., RECHMANN S., RIEGER M., RINKE M., ROSE M.,
RA SCHARFE M., SCHERENS B., SCHOLLER P., SCHWAGER C., SCHWARZ S.,
RA UNDERWOOD A.P., URBESTARAU L.A., VANDENBOL M., VERHAASSEL P.,
RA VIERENDELS F., VOET M., VOLCKAERT G., VOSS H., WAMUTT R., WEDLER E.,
RA WEDLER H., ZIMMERMANN F.K., ZOLLNER A., HANI J., HOHEISEL J.D.;
RL NATURE 387:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C (AB972);
RA DG 2.;
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C (AB972);
RA WATERSTON R.;
RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C (AB972);
RA CHERRY J.M.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U20618; G662136; -.
SQ SEQUENCE 884 AA; 102662 MW; 1A94320F CRC32;

Query Match 6.7%; Score 148; DB 3; Length 884;
Best Local Similarity 26.2%; Pred. No. 4,10e-08;
Matches 68; Conservative 67; Mismatches 93; Indels 32; Gaps 25;

```







# WIPROTECH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution Rights by Oxford Molecular Ltd

Matchup protein - protein database search, using Smith-Waterman algorithm

Run on: FRI Dec 18 18:50:42 1998. Matchup files 23.83 updates/sec  
672,065 MILLION calls updates/sec

Tabular output not generated.

Title: US-08-951-733-14  
Description: 14 parts 15 parts 16 parts 17 parts 18 parts 19 parts 20 parts 21 parts 22 parts 23 parts 24 parts 25 parts 26 parts 27 parts 28 parts 29 parts  
Perfect Score: 949  
Sequence: 1 HASCONC/ALTRATE/LPAT.....PYEDALOGIA/POH/PHML 949

Scoring table: TABLE uninterpretable

Search: 115921 seqs, 1618060 residues

Post-processing: Matching files 1000 summaries

Database: a-geneeq2  
1 part 2 parts 3 parts 4 parts 5 parts 6 parts 7 parts 8 parts 9 parts 10 parts 11 parts 12 parts 13 parts 14 parts 15 parts 16 parts 17 parts 18 parts 19 parts 20 parts 21 parts 22 parts 23 parts 24 parts 25 parts 26 parts 27 parts 28 parts 29 parts

Statistics: Mean 3.259; Variance 0.721; scale 4.521

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being plotted, and is derived by analysis of the total score distribution.

## SUMMARIES

Mon Dec 21 09:28:22 1998

US-08-951-733-14.rag

Page 3

Mon Dec 21 09:28:22 1998

US-08-951-733-14.rag

Page 4

Mon Dec 21 09:28:22 1998

US-08-951-733-14.rag

Page 2

Result No.	Score	Match	Query Length	DB ID	Description	Prod. No.
1	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
2	100	0.7	146	W5531	Sequence of human b5	3.80e+01
3	100	0.7	146	W5531	Reg. region of human b5	3.80e+01
4	100	0.7	146	W5531	Sugar biosynthesis	3.80e+01
5	100	0.7	146	W5531	Proteinase inhibitor	3.80e+01
6	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
7	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
8	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
9	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
10	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
11	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
12	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
13	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
14	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
15	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
16	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
17	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
18	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
19	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
20	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
21	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
22	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
23	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
24	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
25	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
26	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
27	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
28	100	0.7	146	W5531	Human cytochrome b5	3.80e+01
29	100	0.7	146	W5531	Human cytochrome b5	3.80e+01

48	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
49	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
50	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
51	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
52	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
53	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
54	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
55	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
56	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
57	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
58	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
59	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
60	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
61	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
62	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
63	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
64	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
65	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
66	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
67	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
68	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
69	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
70	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
71	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
72	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
73	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
74	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
75	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
76	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
77	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
78	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
79	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
80	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
81	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
82	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
83	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
84	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
85	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
86	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
87	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
88	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
89	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
90	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
91	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
92	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
93	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
94	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
95	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
96	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
97	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
98	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
99	0.6	12.23	W08748	Human cytochrome b5	4.71e+02
100	0.6	12.23	W08748	Human cytochrome b5	4.71e+02











DR WPI: 87-251066/72.  
 PT New gene involved in sugar biosynthesis and attachment - used to generate polypeptide anticonvulsant etc. with altered pattern of glycosylation.  
 PS Disclosure: Fig 4a: 85pp. English.  
 CC EryB1 (M1376) is an enzyme involved in the biosynthesis of the sugar L-mycarose. It is one of 10 enzymes (see also M1374-35 and M1373-41) predicted to be involved in D-desosamine or L-mycarose biosynthesis. The gene is located on chromosome 10. Novel clusters (M1376 and M1378) of saccharopolypolymer erythrose. Novel CC glycosylation-modified polypeptides are produced by selectively CC altering, inactivating or augmenting the ery and/or eryc genes encoding thiazurine biosynthesis enzymes and introducing them into CC cells expressing microsomes.  
 SQ Sequence 333 AA.  
 Query Match 0.7% Score 7: DB 23: Length 333:  
 Best Local Similarity 100.0% Pred. No. 3.80e+01:  
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 DB 15 aligner 21  
 QY 637 ALSTNC 643

RESULT 5  
 ID M45741 standard: Protein: 334 AA.  
 AC M45741 standard: (first entry)  
 DT 05-APR-1994 (first entry)  
 DE Myoinositol dehydrogenase: production: engineering:  
 KW Human: myoinositol dehydrogenase: production: engineering:  
 OS Homo sapiens.  
 PN M06007158-A.  
 PD 18-JAN-1994.  
 PP 23-JUN-1993: 164548.  
 PR 05-APR-1994: US-239431.  
 RA (AAM) : AAM: CHM 2ND CO LTD.  
 DR WPI: 94-05323/07.  
 N-PSDB: Q53708.  
 PT Novel E. coli transformant containing myoinositol dehydrogenase  
 CC dehydrogenase  
 CC Clon 3: Page 8-9: 13pp. Japanese.  
 CC The sequence (Q53708) shows a recombinant plasmid which encodes a  
 CC myoinositol dehydrogenase. The plasmid can be transformed into a  
 CC cell. The host thus producing myoinositol dehydrogenase quickly and  
 CC easily.  
 SQ Sequence 334 AA:

CC Cytokine response protein: CR3; Interleukin-3; IL-3;  
 KW Interleukin-3; Interleukin-3; Interleukin-3; Interleukin-3;  
 OS Homo sapiens.  
 PN M06007158-A.  
 PD 18-JAN-1994.  
 PP 23-JUN-1993: 164548.  
 PR 05-APR-1994: US-239431.  
 RA (AAM) : AAM: CHM 2ND CO LTD.  
 DR WPI: 94-05323/07.  
 N-PSDB: Q53708.  
 PT Novel E. coli transformant containing myoinositol dehydrogenase  
 CC dehydrogenase  
 CC Clon 3: Page 8-9: 13pp. Japanese.  
 CC The sequence (Q53708) shows a recombinant plasmid which encodes a  
 CC myoinositol dehydrogenase. The plasmid can be transformed into a  
 CC cell. The host thus producing myoinositol dehydrogenase quickly and  
 CC easily.  
 SQ Sequence 334 AA:

Query Match 0.7% Score 7: DB 20: Length 358:  
 Best Local Similarity 100.0% Pred. No. 3.80e+01:  
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 DB 245 aligner 251  
 QY 341 POARRG 247

RESULT 8  
 ID W26189 standard: Protein: 375 AA.  
 AC W26189 standard: (first entry)  
 DT 03-DEC-1997 (first entry)  
 DE Human kidney inward rectifier K channel 3 (KIRK-3).  
 KW Human: kidney inward rectifier potassium ion channel: human.  
 OS Homo sapiens.  
 PN M0712974-A2.

Query Match 0.7% Score 7: DB 9: Length 334:  
 Best Local Similarity 100.0% Pred. No. 3.80e+01:  
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 DB 98 aligner 104  
 QY 304 ATSLDA 310

RESULT 6  
 ID M44246 standard: Protein: 358 AA.  
 AC M44246 standard: (first entry)  
 DT 15-MAR-1998 (first entry)  
 DE Human HP4 prostaglandin receptor.  
 KW Human: HP4 prostaglandin receptor: adenylate cyclase: drug screening:  
 OS Homo sapiens.  
 PN M0516815-A.  
 PD 28-FEB-1998: 834706.  
 PP 05-MAR-1994: US-239431.  
 PR 05-MAR-1994: US-239431.  
 RA (AAM) : AAM: CHM 2ND CO LTD.  
 DR WPI: 96-14807/3.  
 N-PSDB: V12457.  
 PT DNA encoding human HP4 prostaglandin receptor - useful for drug  
 CC screening.  
 CC The present sequence represents human HP4 prostaglandin receptor.  
 CC Transfected cells, containing an HP4 prostaglandin receptor  
 CC expression vector, can be used to screen for substances that  
 CC bind to the HP4 receptor, for substances that inhibit ligand  
 CC binding, for substances that increase or decrease HP4 receptor  
 CC (based on increased cAMP production in cells pretreated with a  
 CC phosphodiesterase inhibitor).  
 SQ Sequence 358 AA.  
 Query Match 0.7% Score 7: DB 29: Length 358:  
 Best Local Similarity 100.0% Pred. No. 3.80e+01:  
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 DB 245 aligner 251  
 QY 241 POARRG 247

RESULT 7  
 ID M08115 standard: Protein: 358 AA.  
 AC M08115 standard: (first entry)  
 DT 11-MAR-1997 (first entry)  
 DE Human cytokine response protein CR3.

PD 10-APR-1997.  
 PR 09-SEP-1996: U14134.  
 PR 15-SEP-1995: US-003819.  
 RA (PMA) : PHARMACIA & UPJOHN CO.  
 PN M0712974-A2.  
 PD 18-JAN-1994.  
 PP 23-JUN-1993: 164548.  
 PR 05-APR-1994: US-239431.  
 RA (AAM) : AAM: CHM 2ND CO LTD.  
 DR WPI: 94-05323/07.  
 N-PSDB: Q53708.  
 PT Novel E. coli transformant containing myoinositol dehydrogenase  
 CC dehydrogenase  
 CC Clon 3: Page 8-9: 13pp. Japanese.  
 CC The sequence (Q53708) shows a recombinant plasmid which encodes a  
 CC myoinositol dehydrogenase. The plasmid can be transformed into a  
 CC cell. The host thus producing myoinositol dehydrogenase quickly and  
 CC easily.  
 SQ Sequence 334 AA:

Query Match 0.7% Score 7: DB 34: Length 375:  
 Best Local Similarity 100.0% Pred. No. 3.80e+01:  
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 DB 260 aligner 266  
 QY 804 ATSLDA 810

RESULT 9  
 ID M80288 standard: Protein: 395 AA.  
 AC M80288 standard: (first entry)  
 DT 21-DEC-1995 (first entry)  
 DE Galactose galactose 4-epimerase: galactose galactose 4-epimerase:  
 KW Human: galactose galactose 4-epimerase: galactose galactose 4-epimerase:  
 OS Homo sapiens.  
 PN M0543530-A.  
 PD 25-JUL-1993.  
 PP 28-FEB-1998: 834706.  
 PR 10-APR-1997: US-009418.  
 PR 29-APR-1993: US-692769.  
 PR 19-OCT-1992: US-967949.  
 PR 10-JUL-1991: US-099979.  
 RA (AAM) : AAM: CHM 2ND CO LTD.  
 DR WPI: 96-14807/3.  
 N-PSDB: V12457.  
 PT DNA encoding human HP4 prostaglandin receptor - useful for drug  
 CC screening.  
 CC The present sequence represents human HP4 prostaglandin receptor.  
 CC Transfected cells, containing an HP4 prostaglandin receptor  
 CC expression vector, can be used to screen for substances that  
 CC bind to the HP4 receptor, for substances that inhibit ligand  
 CC binding, for substances that increase or decrease HP4 receptor  
 CC (based on increased cAMP production in cells pretreated with a  
 CC phosphodiesterase inhibitor).  
 SQ Sequence 358 AA.

[illegible]

Mon Dec 21 09:28:22 1998

US-08-951-733-14.rag

Page 31

01 Disordered. Page 55-61: 81pp. English.  
 02 CC The gall gene product (p70227) is known as galactose-1-phosphate  
 03 CC uridylyltransferase (transferase), the gall gene product (p70225) is  
 04 CC known as uridine diphosphoglucose-4-epimerase (epimerase), and  
 05 CC the gall gene product (p70226) is known as galactose-1-phosphate  
 06 CC epimerase. The gall gene product (p70227) is known as galactose-1-phosphate  
 07 CC foreign DNA in a transformed micro-organism.  
 08 CC Sequence 395 AA:  
 09  
 10 CC Query Match 0.78; Score 71; Length 395;  
 11 CC Best Local Similarity 100.0%; Pos. No. 1; Gap 0;  
 12 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 13 CC 07 260 PROBAB 266  
 14  
 15 CC RESULT 13  
 16 CC IT 04232 standard; protein: 198 AA.  
 17 CC AC W24232;  
 18 CC DE 18-FEB-1989 (first entry)  
 19 CC DT Aminoacyl deaminase/aspartate aminotransferase KCA/AA.  
 20 CC OS Aminoacyl deaminase/aspartate KCA/AA, chiral compound.  
 21 CC OS Aminoacyl deaminase/aspartate KCA/AA, chiral compound.  
 22 CC MO3793187-AL  
 23  
 24 CC PD 14-AUG-1987;  
 25 CC PF 21-JUN-1987; 000194;  
 26 CC PR 08-MAY-1986; 05-595120;  
 27 CC PR 08-MAY-1986; 05-595120;  
 28 CC PA (RECD) RECOMBINANT BLOOMALYSIS INC.  
 29 CC P1 Swenson RV, Warren PV;  
 30 CC MPI: 97-415134/38.  
 31  
 32 CC DR R-PS08; 778776;  
 33 CC PT used for producing enzymes, and hybridization probes for a cDNA or  
 34 CC genomic library  
 35 CC CC  
 36 CC CC  
 37 CC CC  
 38 CC CC  
 39 CC CC  
 40 CC CC  
 41 CC CC  
 42 CC CC  
 43 CC CC  
 44 CC CC  
 45 CC CC  
 46 CC CC  
 47 CC CC  
 48 CC CC  
 49 CC CC  
 50 CC CC  
 51 CC CC  
 52 CC CC  
 53 CC CC  
 54 CC CC  
 55 CC CC  
 56 CC CC  
 57 CC CC  
 58 CC CC  
 59 CC CC  
 60 CC CC  
 61 CC CC  
 62 CC CC  
 63 CC CC  
 64 CC CC  
 65 CC CC  
 66 CC CC  
 67 CC CC  
 68 CC CC  
 69 CC CC  
 70 CC CC  
 71 CC CC  
 72 CC CC  
 73 CC CC  
 74 CC CC  
 75 CC CC  
 76 CC CC  
 77 CC CC  
 78 CC CC  
 79 CC CC  
 80 CC CC  
 81 CC CC  
 82 CC CC  
 83 CC CC  
 84 CC CC  
 85 CC CC  
 86 CC CC  
 87 CC CC  
 88 CC CC  
 89 CC CC  
 90 CC CC  
 91 CC CC  
 92 CC CC  
 93 CC CC  
 94 CC CC  
 95 CC CC  
 96 CC CC  
 97 CC CC  
 98 CC CC  
 99 CC CC  
 100 CC CC  
 101 CC CC  
 102 CC CC  
 103 CC CC  
 104 CC CC  
 105 CC CC  
 106 CC CC  
 107 CC CC  
 108 CC CC  
 109 CC CC  
 110 CC CC  
 111 CC CC  
 112 CC CC  
 113 CC CC  
 114 CC CC  
 115 CC CC  
 116 CC CC  
 117 CC CC  
 118 CC CC  
 119 CC CC  
 120 CC CC  
 121 CC CC  
 122 CC CC  
 123 CC CC  
 124 CC CC  
 125 CC CC  
 126 CC CC  
 127 CC CC  
 128 CC CC  
 129 CC CC  
 130 CC CC  
 131 CC CC  
 132 CC CC  
 133 CC CC  
 134 CC CC  
 135 CC CC  
 136 CC CC  
 137 CC CC  
 138 CC CC  
 139 CC CC  
 140 CC CC  
 141 CC CC  
 142 CC CC  
 143 CC CC  
 144 CC CC  
 145 CC CC  
 146 CC CC  
 147 CC CC  
 148 CC CC  
 149 CC CC  
 150 CC CC  
 151 CC CC  
 152 CC CC  
 153 CC CC  
 154 CC CC  
 155 CC CC  
 156 CC CC  
 157 CC CC  
 158 CC CC  
 159 CC CC  
 160 CC CC  
 161 CC CC  
 162 CC CC  
 163 CC CC  
 164 CC CC  
 165 CC CC  
 166 CC CC  
 167 CC CC  
 168 CC CC  
 169 CC CC  
 170 CC CC  
 171 CC CC  
 172 CC CC  
 173 CC CC  
 174 CC CC  
 175 CC CC  
 176 CC CC  
 177 CC CC  
 178 CC CC  
 179 CC CC  
 180 CC CC  
 181 CC CC  
 182 CC CC  
 183 CC CC  
 184 CC CC  
 185 CC CC  
 186 CC CC  
 187 CC CC  
 188 CC CC  
 189 CC CC  
 190 CC CC  
 191 CC CC  
 192 CC CC  
 193 CC CC  
 194 CC CC  
 195 CC CC  
 196 CC CC  
 197 CC CC  
 198 CC CC  
 199 CC CC  
 200 CC CC  
 201 CC CC  
 202 CC CC  
 203 CC CC  
 204 CC CC  
 205 CC CC  
 206 CC CC  
 207 CC CC  
 208 CC CC  
 209 CC CC  
 210 CC CC  
 211 CC CC  
 212 CC CC  
 213 CC CC  
 214 CC CC  
 215 CC CC  
 216 CC CC  
 217 CC CC  
 218 CC CC  
 219 CC CC  
 220 CC CC  
 221 CC CC  
 222 CC CC  
 223 CC CC  
 224 CC CC  
 225 CC CC  
 226 CC CC  
 227 CC CC  
 228 CC CC  
 229 CC CC  
 230 CC CC  
 231 CC CC  
 232 CC CC  
 233 CC CC  
 234 CC CC  
 235 CC CC  
 236 CC CC  
 237 CC CC  
 238 CC CC  
 239 CC CC  
 240 CC CC  
 241 CC CC  
 242 CC CC  
 243 CC CC  
 244 CC CC  
 245 CC CC  
 246 CC CC  
 247 CC CC  
 248 CC CC  
 249 CC CC  
 250 CC CC  
 251 CC CC  
 252 CC CC  
 253 CC CC  
 254 CC CC  
 255 CC CC  
 256 CC CC  
 257 CC CC  
 258 CC CC  
 259 CC CC  
 260 CC CC  
 261 CC CC  
 262 CC CC  
 263 CC CC  
 264 CC CC  
 265 CC CC  
 266 CC CC  
 267 CC CC  
 268 CC CC  
 269 CC CC  
 270 CC CC  
 271 CC CC  
 272 CC CC  
 273 CC CC  
 274 CC CC  
 275 CC CC  
 276 CC CC  
 277 CC CC  
 278 CC CC  
 279 CC CC  
 280 CC CC  
 281 CC CC  
 282 CC CC  
 283 CC CC  
 284 CC CC  
 285 CC CC  
 286 CC CC  
 287 CC CC  
 288 CC CC  
 289 CC CC  
 290 CC CC  
 291 CC CC  
 292 CC CC  
 293 CC CC  
 294 CC CC  
 295 CC CC  
 296 CC CC  
 297 CC CC  
 298 CC CC  
 299 CC CC  
 300 CC CC  
 301 CC CC  
 302 CC CC  
 303 CC CC  
 304 CC CC  
 305 CC CC  
 306 CC CC  
 307 CC CC  
 308 CC CC  
 309 CC CC  
 310 CC CC  
 311 CC CC  
 312 CC CC  
 313 CC CC  
 314 CC CC  
 315 CC CC  
 316 CC CC  
 317 CC CC  
 318 CC CC  
 319 CC CC  
 320 CC CC  
 321 CC CC  
 322 CC CC  
 323 CC CC  
 324 CC CC  
 325 CC CC  
 326 CC CC  
 327 CC CC  
 328 CC CC  
 329 CC CC  
 330 CC CC  
 331 CC CC  
 332 CC CC  
 333 CC CC  
 334 CC CC  
 335 CC CC  
 336 CC CC  
 337 CC CC  
 338 CC CC  
 339 CC CC  
 340 CC CC  
 341 CC CC  
 342 CC CC  
 343 CC CC  
 344 CC CC  
 345 CC CC  
 346 CC CC  
 347 CC CC  
 348 CC CC  
 349 CC CC  
 350 CC CC  
 351 CC CC  
 352 CC CC  
 353 CC CC  
 354 CC CC  
 355 CC CC  
 356 CC CC  
 357 CC CC  
 358 CC CC  
 359 CC CC  
 360 CC CC  
 361 CC CC  
 362 CC CC  
 363 CC CC  
 364 CC CC  
 365 CC CC  
 366 CC CC  
 367 CC CC  
 368 CC CC  
 369 CC CC  
 370 CC CC  
 371 CC CC  
 372 CC CC  
 373 CC CC  
 374 CC CC

```

PF 07-SEP-1993.
PD 28-FEB-1986. 834706.
PR 28-FEB-1986 US-834706.
PR 28-FEB-1986 US-692769.
PR 29-APR-1991 US-692769.
PR 27-OCT-1993 US-967949.
PA (SMIT ) SMTIMLINE BECKMAN CORP.
D1 WPM 93.75523479 HZ. Fornwald JA, Schmidt PJ;
DR N-P5081 Q48439.
P7 Streptococcus, galactose 5'AI operon sequence - Including P1 and P2
P7 promoter, galK, galE and galT genes, for enabling transformed
CC strain 2; Page 21-25, 26pp. English.
CC Sequence Q48439 is the complete DNA sequence of the S.lividans gal
CC operon. M1d-type promoter p1 lies upstream of three open readings
CC deduced from the ORFs have some homology with the respective
CC enzymes (galactose-1-phosphate uridylyltransferase, UDP-1-epimerase
CC and galactokinase) from E.coli and Y.cerevisiae.
CC Sequence R41535 and R41539.
SC Sequence 395 AAs.

Query Match      0.7% Score 7; DB 8; Length 395;
Bect. Koch Similarity 100.0%; Pred. No. 3,84e-01;
Matches 77; Conservative 0; Mismatches 0; Insets 0; Gaps 0
Db   375 Synonym 181
Cy   260 PROBAB 166
|||||||

RESULT:
P1 28-FEB-1993 standard; Protein: 395 AAs.
AD P70216.
DE 28-FEB-1993 (first entry)
DE Sequence encoded by s.lividans gal operon gene.
DE Galactose metabolism; promoter; expression vector; gal operon.
PI 28-FEB-1986 S.lividans strain 1936.
PI 27-SEP-1993.
PD 02-SEP-1987.
PF 28-FEB-1987. 870026.
PR 10-JAN-1987 US-009419.
PR 10-JAN-1987 US-009419.
PA (SMIT ) SMTIMLINE BECKMAN CORP.
P1 Adams CW, Fornwald JA, Brauner M, Schmidt PJ;
DR N-P5081 Q48412755.
P7 New recombinant DNA consg. streptococcus gal operon - or its
P7 promoter and gene fragments, and transformed cell able to express
P7 foreign protein;and to metabolize galactose
```

MOD Dec 21 09:28:22 1998

US-08-951-733-14.r09

Page 31

```

50 Sequence 390 AA;
Query Match 0.74; Score 7; DB 26; Length 388;
Best Local Similarity 100.0%; Pred. No. 3,80e+01;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 157 IAAATVP 443
|||||
QY 437 IAAATVP 443

RESULT 13
ID R88407 standard; Protein 429 AA.
AC R88407.
DT 05-JUL-1996 (first entry)
KW Trichoderma harzianum beta-(1,6)-endoglucanase; Aspergillus oryzae;
KW beta-glucan degradation; pustulanase.
OS Trichoderma harzianum.
PN T05512314.1.
PI 11-MAY-1993; D0189.
PR 11-MAY-1993; D0189.
PA (MOV) / MOV0-NORDISK AS.
P1 Nordisk Lactose A/S.
PI 11-MAY-1993; D0189.
DR WP1: 96-010922/01.
PT DNA encoding beta-1,6-endoglucanase from Trichoderma harzianum -
P1 useful, e.g., in prep. of yeast extracts, as antifungal agent, in
P5 strain 15, Page 35; 48pp: English.
P7 Glus 11, Page 35; 48pp: English.
CC Trichoderma harzianum (CBS 243.7) beta-(1,6)-endoglucanase may 50,000,
CC produced recombinantly, and has a molecular weight of about 50,000,
CC in apparent isoelectric point of 5.6, an optimal temperature of
CC activity of about 100 °/deg. The enzyme's compositions enriched
CC in it, are used to modify or degrade beta-glucans, particularly for
CC rupturing or lysing cell walls of microorganisms thereby enabling
CC recovery of desirable products produced by the microorganisms
CC and/or to remove undesirable components from the products.
CC Yeast extracts, use in wine and press juice making, as fungicides,
CC to remove excess dye from textiles, to remove moulds on coatings or
CC biofilms from surfaces, for cleaning dentures and removing plaque,
CC and in the extraction of mannoproteins from microbial cell walls.
CC sequence 429 AA.
Query Match 0.74; Score 7; DB 17; Length 429;
Best Local Similarity 100.0%; Pred. No. 3,80e+01;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

DB 160 AWWHSA 166  
QY 767 AWWHSA 773

RESULT 14  
ID R87508 standard: Protein: 493 AA.  
AC R87508: (first entry)  
DT 15-OCT-1995: (first entry)  
DE Jojoba: fatty-acyl-reductase; enzyme: transgenic plant; wax ester:  
NM Brassica.  
OS Simmondsia chinensis.  
PM 07-DEC-1995: A2.  
PR 01-JUN-1995: US-251464.  
PA (CALJ) CALGENE INC.  
PI Lardilabai RD, Lesmer MW, Metz JG.  
DR N-PDB: T06149/70.  
PT Synthetic Jojoba fatty acyl reductase gene having reduced AT content  
- used in the production of wax ester in plant cells not naturally  
producing wax ester.  
PS Disclosure: Fig 11-7: 7pp. English.  
CC This wild-type fatty-acyl-reductase may be reengineered such  
that it has an AT content of less than 5% (especially less than  
5.1%) (T06150). The sequence may be expressed in Brassica sp., used  
to produce new oil and wax compositions containing a 4:2 wax ester  
as a predominant component.  
Sequence 493 AA:

Query Match  
Best Local Similarity 100.0%; Pred. No. 3,800+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 426 111r1v4 432  
QY 884 LILRLVD 890

RESULT 15  
ID R26898 standard: Protein: 493 AA.  
AC R26898: (first entry)  
DT 15-FEB-1993: (first entry)  
DE Jojoba: fatty-acyl reductase.  
NM Jojoba: fatty-acyl reductase.  
OS Jojoba embryo.  
PM 03-SEP-1993: A2.

CC new oil and wax compositions containing a 4:2 wax ester as a  
CC predominant component.  
Sequence 493 AA:

Query Match  
Best Local Similarity 100.0%; Pred. No. 3,800+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 426 111r1v4 432  
QY 884 LILRLVD 890

RESULT 17  
ID R7163 standard: Protein: 493 AA.  
AC R7163: (first entry)  
DT 04-DEC-1995: (first entry)  
DE Jojoba: fatty-acyl reductase.  
NM Very long chain fatty acid: acyl-CoA: fatty acyl reductase;  
NM wax synthase.  
PM 08-JUN-1995: A2.  
PR 30-NOV-1994: US-186602.  
PA (CALJ) CALGENE INC.  
PI Lardilabai RD, Lesmer MW, Metz JG.  
DR N-PDB: T06149/70.  
PT DNA construct expressing Jojoba wax synthase and transgenically  
produced wax ester in plant cells not naturally producing wax ester.  
PS Disclosure: Fig 11: 14pp. English.  
CC Application of Jojoba wax synthase protein is described in PCT  
application of Jojoba: fatty-acyl reductase and wax synthase  
proteins. Jojoba wax synthase protein having wax synthase activity isolated.  
CC The wax synthase activity is solubilized, and the wax synthase  
protein is further purified. The source material is the outer shells  
and seed coats of Jojoba embryos.  
Sequence 493 AA:

Query Match  
Best Local Similarity 100.0%; Pred. No. 3,800+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 426 111r1v4 432  
QY 884 LILRLVD 890

PR 21-FEB-1993: US-055975.  
PR 27-SEP-1991: US-167251.  
PR 20-NOV-1991: US-795256.  
PI Metz JG, Pollard NR.  
DR WPI: 92-31670/38.  
DR N-PDB: Q28369.  
PT Plant seed e.g. Brassica, fatty acyl reductase enzymes - produced  
in transgenic plant cells, naturally convert fatty acyl  
substrates to fatty alcohols (e.g.).  
PS Disclosure: Fig 11-7: 8pp. English.  
CC A Jojoba cDNA library was screened using PCR primers designed from  
the Jojoba fatty-acyl reductase cDNA (US-055975) and a Jojoba  
fatty-acyl reductase protein was obtained and resequenced. The sequence may be  
used in the prodn. of PAR in recombinant DNA technology.  
Sequence 493 AA:

Query Match  
Best Local Similarity 100.0%; Pred. No. 3,800+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 426 111r1v4 432  
QY 884 LILRLVD 890

RESULT 16  
ID R87509 standard: Protein: 493 AA.  
AC R87509: (first entry)  
DT 15-OCT-1995: (first entry)  
DE Resynthesized Jojoba fatty-acyl-reductase  
NM Jojoba: fatty-acyl-reductase; enzyme: transgenic plant; wax ester:  
OS Simmondsia chinensis.  
PM 07-DEC-1995: A2.  
PR 01-JUN-1995: US-251464.  
PA (CALJ) CALGENE INC.  
PI Lardilabai RD, Lesmer MW, Metz JG.  
DR N-PDB: T06149/70.  
PT Synthetic Jojoba fatty acyl reductase gene having reduced AT content  
- used in the production of wax ester in plant cells not naturally  
producing wax ester.  
PS Disclosure: Fig 11-7: 7pp. English.  
CC This wild-type fatty-acyl-reductase has had its AT content  
reduced from 57.5% present in the wild-type (R87508) to about  
5.1%. The sequence may be expressed in Brassica sp. transgenic  
plants so they produce a wax ester. It may also be used to produce

ID R79934 standard: Protein: 493 AA.  
AC R79934: (first entry)  
DT 16-JAN-1996: (first entry)  
DE Jojoba: fatty-acyl-reductase.  
NM Fatty-acyl reductase; wax synthase; Jojoba: oilseed;  
NM Jojoba: fatty-acyl reductase; wax synthase; wax ester;  
NM transgenic plant; crop improvement: Brassica.  
OS Simmondsia chinensis.  
PM 05-4-1997-A.  
PR 28-NOV-1995: 765256.  
PR 20-NOV-1991: US-786355.  
PR 21-NOV-1992: US-913411.  
PR 13-NOV-1993: WO-009863.  
PR 20-NOV-1993: WO-009863.  
PM 07-DEC-1995: A2.  
PI Lardilabai RD, Lesmer MW, Metz JG.  
DR WPI: 95-31089/40.  
DT N-PDB: T04125.  
PT DNA construct expressing Jojoba wax synthase and transgenically  
produced wax ester in plant cells not naturally producing wax ester for use in  
PT pharmaceuticals and cosmetics, etc.  
PS Disclosure: column 43-46: 50pp. English.  
CC A Jojoba fatty-acyl-reductase cDNA was used in the construction of  
a Jojoba fatty-acyl reductase protein. The protein was solubilized  
and the reductase activity was assayed. The reductase is added to a heterologous  
wax synthase are used for wax ester prodn.  
Sequence 493 AA:

Query Match  
Best Local Similarity 100.0%; Pred. No. 3,800+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 426 111r1v4 432  
QY 884 LILRLVD 890

RESULT 18  
ID R7163 standard: Protein: 493 AA.  
AC R7163: (first entry)  
DT 14-NOV-1995: (first entry)  
DE Jojoba: fatty-acyl-CoA reductase.  
NM Jojoba: fatty-acyl-CoA reductase; fatty acyl: fatty alcohol.  
OS Simmondsia chinensis.  
PM 06-DEC-1994: 656975.  
PR 21-FEB-1991: US-656975.  
PR 27-SEP-1991: US-767251.  
PR 20-NOV-1991: US-795256.















CC cultured to produce the protein. The alpha-glucuronidase protein is used  
CC to improve dough and baked products. The protein can also be used to  
CC treat cellulose pulp to remove at least some of the glucuronic acid  
CC residue present in it, as a silaging additive and to degrade plant  
CC matter generally. When added to animal feed, particularly the protein  
CC increases feed utilization and feed conversion rate, particularly by  
CC increasing bioavailability of metal ions.

50 Query Match 0.64: Score 6; DN 28; Length 35;

Best Local Similarity 85.7%: Pred. No. 4, 7, 1e+02;

Matches 6: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 5 align 11

07 857 SILESTL 863

RESULT 50  
ID W20450 standard: Protein: 45 AA.

AC W20450: 11-JUN-1997 (first entry)

DT 11-JUN-1997 (first entry)

RD 11-JUN-1997 (first entry)

RE 11-JUN-1997 (first entry)

RV 11-JUN-1997 (first entry)

RW 11-JUN-1997 (first entry)

RM 11-JUN-1997 (first entry)

RS 11-JUN-1997 (first entry)

RT 11-JUN-1997 (first entry)

RU 11-JUN-1997 (first entry)

RV 11-JUN-1997 (first entry)

RW 11-JUN-1997 (first entry)

RM 11-JUN-1997 (first entry)

RS 11-JUN-1997 (first entry)

RT 11-JUN-1997 (first entry)

RU 11-JUN-1997 (first entry)

RV 11-JUN-1997 (first entry)

RW 11-JUN-1997 (first entry)

RM 11-JUN-1997 (first entry)

RS 11-JUN-1997 (first entry)

RT 11-JUN-1997 (first entry)

RU 11-JUN-1997 (first entry)

RV 11-JUN-1997 (first entry)

RW 11-JUN-1997 (first entry)

RM 11-JUN-1997 (first entry)

RS 11-JUN-1997 (first entry)

RT 11-JUN-1997 (first entry)

RU 11-JUN-1997 (first entry)

RV 11-JUN-1997 (first entry)

RW 11-JUN-1997 (first entry)

RM 11-JUN-1997 (first entry)

RS 11-JUN-1997 (first entry)

RT 11-JUN-1997 (first entry)

RU 11-JUN-1997 (first entry)

RV 11-JUN-1997 (first entry)

RW 11-JUN-1997 (first entry)

RM 11-JUN-1997 (first entry)

RS 11-JUN-1997 (first entry)

RT 11-JUN-1997 (first entry)

RU 11-JUN-1997 (first entry)

RV 11-JUN-1997 (first entry)

RW 11-JUN-1997 (first entry)

RM 11-JUN-1997 (first entry)

RS 11-JUN-1997 (first entry)

RT 11-JUN-1997 (first entry)

RU 11-JUN-1997 (first entry)

RV 11-JUN-1997 (first entry)

RW 11-JUN-1997 (first entry)

RM 11-JUN-1997 (first entry)

RS 11-JUN-1997 (first entry)

CC acid sequences predicted from various ORF were analyzed for significant  
CC homology to other known or exported membrane proteins. Having identified  
CC and determined the sequences of interest, particular regions can be  
CC isolated from R. pylori for application for recombinant polypeptide  
CC vaccines in R. coli hosts.

50 Query Match 0.64: Score 6; DN 22; Length 45;

Best Local Similarity 100%: Pred. No. 4, 7, 1e+02;

Matches 6: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 8 align 13

07 682 MBSVL 687

Search completed: Fri Dec 18 18:54:12 1998  
Job time : 210 secs.















```

CAMP-dependent protein kinase II from porcine skeletal muscle
across-reference NCID 8006605
across-molecule-type protein
across-label POT
across-database A15740
REFERENCE
1. Blot, Chen, (1980) 235:8483-8488
across-author Blot, Chen
across-journal J Biol Chem
across-volume 235
across-issue 8
across-first-page 8483
across-last-page 8488
across-title Covalent modification of an adenosine 3',5'-monophosphate binding site of the regulatory subunit of cAMP-dependent protein kinase II with 8-azidoadenosine
across-keywords
across-reference NCID 8106982
accession A15740
molecule-type protein
across-molecule-type protein
across-label PKA
COMMENT
This catalytic form of the enzyme is composed of two regulatory chains and two catalytic chains. Activation by cAMP produces four active catalytic monomers and a regulatory dimer that binds four cAMP molecules.
COMMENT
Forskolin, a cyclic nucleotide regulatory chain are found: 1-alpha, 8-1beta, 11-alpha, and 11-beta. Their expression varies among tissues and is in some cases constitutive and in others inducible.
COMMENT
Type II regulatory chains are phosphorylated and activated by phosphorylation by other kinases is unclear.
COMMENT
Type II regulatory chains mediate membrane association by binding to anchoring proteins, including the MARCKS kinase. The phosphorylation of the regulatory chains is essential for this binding and to dimerization. 30 residues are essential for this binding and to dimerization.
CLASSIFICATION
superfamily cAMP-dependent protein kinase regulatory chain; cAMP receptor protein cyclic nucleotide-binding domain
KEYWORDS
cAMP binding; duplication; heterotrimer; homodimer; phosphoprotein; phosphotransferase
SYNOPSIS
domain cAMP receptor protein cyclic nucleotide-binding domain (cAMP-dependent); label CAP
length 169; checksum 355
SUMMARY
Query Match 0.88; Score 8; DB: 2; length 169;
Accession Match 0.88; Score 22.0; length 169;
Machine Matched 0; Mismatches 0; Indels 0; Gaps 0;
QY 461 EEDTDR 468
RESULT 5

```





ALTERNATE\_NAMES myosin alkali light chain  
 ORIGINISM formal\_name Mus musculus scconcom\_name mouse  
 DATE 01-Dec-1998 sequence\_revision 01-Dec-1998 text\_change  
 01-Dec-1998  
 ACCSSIONS S01345  
 REFERENCE S01344  
 #authors Cohen, A.; Barton, P.J.R.; Robert, B.; Garner, I.; Alonso,  
 #journal J Biol Chem 267(1992) 16:10037-10052  
 #title Promoter analysis of myosin alkali light chain genes  
 #comment expressed in mouse striated muscle.  
 #cross-references MID:89057447  
 #molecule-type DNA  
 #residues 1-52 label COM  
 #accession EMBL:U12972  
 #classification superfamily: calmodulin repeat homology  
 #keywords repeats  
 #summary length 52 checksum 8566

Query Match 0.74: Score 7; DB 2; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 34 APPPER 40  
 QY 264 APPPER 270

RESULT 16  
 ENTRY A49196 type fragment  
 #authors Roselli-Rehman, L.; Robbins, L.S.; Cone, R.D.  
 #journal Endocrinology (1992) 130:1837-1861  
 #title Thyrotropin receptor messenger ribonucleic acid is expressed  
 #cross-references MID:2191870  
 #accession A49196  
 #status preliminary; not compared with conceptual translation  
 #molecule-type nucleic acid  
 #residues 1-103  
 #comment sequence extracted from MCB1 backbone (NCBIP:89464)  
 #note superfamily: glycoprotein hormone receptor; leucine-rich  
 #classification alpha-2-glycoprotein repeat homology

cross-references MID:8604665  
 #accession A23329  
 #molecule-type RNA  
 #residues 1-115 label SHE  
 #cross-references MID:9198919; PID:9198920  
 #classification protein-cytosolic domain homology;  
 antigen cytosolic domain homology;  
 protein-cytosine-phosphatase homology  
 #keywords glycoprotein  
 #summary length 115 checksum 1887

Query Match 0.74: Score 7; DB 2; Length 115;  
 Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 121 POFALD 127  
 QY 121 POFALD 127

RESULT 19  
 ENTRY S77068 type complete  
 #authors Saito, S.; Kozaki, H.; Tanaka, A.; Asanuma, E.;  
 Nakamura, Y.; Miyajima, M.; Hirose, M.; Sugawara, M.;  
 Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsumoto, A.;  
 Muraki, A.; Nakasaki, M.; Haruo, K.; Okumura, S.; Shimo,  
 Y.; Yamauchi, T.; Tanaka, T.; Watanabe, A.; Tanaka, M.;  
 #journal DNA Res. (1998) 5:109-116  
 #title Sequence analysis of the genome of the unclonable  
 #cross-references MID:97061201  
 #accession S77068  
 #status preliminary  
 #molecule-type DNA  
 #residues 1-128 label KAN  
 #cross-references MID:97061201; PID:97061201  
 #classification the nucleotide sequence was submitted to the EMBL Data  
 #summary length 128 molecular-weight 14443 checksum 2173

Query Match 0.74: Score 7; DB 2; Length 128;

length 52 checksum 4696

Query Match 0.74: Score 7; DB 2; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 46 LELVAN 52  
 QY 99 LELVAN 105

RESULT 17  
 ENTRY S14855 type complete  
 #authors Kuznetsov, A.I.; Serasing, S.A.; Tikhonova, L.P.; Kravkov,  
 #journal submitted to the EMBL Data Library, April 1991  
 #description Nucleotide sequence of Kanamycin Polymorphic DNA region  
 #accession S14855  
 #status preliminary  
 #molecule-type DNA  
 #residues 1-111 label KNU  
 #cross-references MID:87762; PID:87764  
 #summary length 111 molecular-weight 12281 checksum 3577

Query Match 0.74: Score 7; DB 2; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 22 TSLART 28  
 QY 640 TSLART 646

RESULT 18  
 ENTRY A23329 type fragment  
 #authors Roselli-Rehman, L.; Robbins, L.S.; Cone, R.D.  
 #journal Endocrinology (1992) 130:1837-1861  
 #title Thyrotropin receptor messenger ribonucleic acid is expressed  
 #cross-references MID:2191870  
 #accession A23329  
 #status preliminary; not compared with conceptual translation  
 #molecule-type nucleic acid  
 #residues 1-103  
 #comment sequence extracted from MCB1 backbone (NCBIP:89464)  
 #note superfamily: glycoprotein hormone receptor; leucine-rich  
 #classification alpha-2-glycoprotein repeat homology

Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 46 AWTGDS 52  
 QY 811 AWTGDS 817

RESULT 20  
 ENTRY A23329 type complete  
 #authors Roselli-Rehman, L.; Robbins, L.S.; Cone, R.D.  
 #journal Endocrinology (1992) 130:1837-1861  
 #title Thyrotropin receptor messenger ribonucleic acid is expressed  
 #cross-references MID:2191870  
 #accession A23329  
 #status preliminary; not compared with conceptual translation  
 #molecule-type nucleic acid  
 #residues 1-103  
 #comment sequence extracted from MCB1 backbone (NCBIP:89464)  
 #note superfamily: glycoprotein hormone receptor; leucine-rich  
 #classification alpha-2-glycoprotein repeat homology

Query Match 0.74: Score 7; DB 2; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 4 LIGASV 10

697 LCA5VL 703

RESULT 21

ENTRY 510280

GENETICS \*type complete

ORGANISM hsp protein - Saccharicola coli

DATE 30-Sep-1993 sequence\_revision 10-Sep-1993 text\_change

ACCESSIONS 510279

REFERENCE 510279

author Roper, D.I.; Pavetti, T.; Cooper, R.A.

journal Mol. Gen. Genet. (1993) 237:241-250

title The *Saccharicola coli* C homoproteolytic degradation control of expression.

accession 510280

molecule\_type DNA

residues 1118

cross-references DB:55655; NID:929654; PID:929655

GENETICS

gene hpcR

REMARKS DNA binding: transcription regulation

SUMMARY length 148 molecule\_weight 17355 checksum 5904

Query Match 0.7% Score 7; DB 2; Length 148;

Best Local Similarity 100.0%; Pred. No. 1.04e+01;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 58 LRP5LTC 64

697 LCA5VL 703

RESULT 22

ENTRY 564794

GENETICS \*type complete

ORGANISM hypothetical protein YLL042c - Yeast (Saccharomyces

DATE 01-Aug-1995 sequence\_revision 24-May-1996 text\_change

ACCESSIONS 564794

REFERENCE 564794

author Medler, H.; Wedler, E.; Schaefer, M.; Wambutt, R.

journal submitted to the Protein Sequence Database, May 1996

title

accession 564794

molecule\_type DNA

residues 1167

cross-references DBL:273147; NID:91560316; PID:91560317;

cross-references DBL:273147; NID:91560316; PID:91560317;

cross-references DBL:273147; NID:91560316; PID:91560317;

697 LCA5VL 703

RESULT 24

ENTRY 505418

GENETICS \*type complete

ORGANISM hsp protein - Pseudomonas aeruginosa

DATE 07-Jul-1997 sequence\_revision 29-Aug-1997 text\_change

ACCESSIONS 505418

REFERENCE 505418

author Hishida, T.; Iwazaki, H.; Ishioke, K.; Shingawa, H.

journal Gene (1996) 182:63-70

title Molecular analysis of the *Pseudomonas aeruginosa* genes, *hsp*, *hsp70* and *hsp90*, involved in processing of homologous protein degradation intermediates.

accession 505418

molecule\_type DNA

residues 1114

cross-references DB:55655; NID:918337; PID:918337; PID:918337;

cross-references DB:55655; NID:918337; PID:918337; PID:918337;

cross-references DB:55655; NID:918337; PID:918337; PID:918337;

GENETICS

gene hsp70

REMARKS

SUMMARY length 174 molecule\_weight 18555 checksum 8754

Query Match 0.7% Score 7; DB 2; Length 174;

Best Local Similarity 100.0%; Pred. No. 1.04e+01;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 164 CARRRG 170

697 LCA5VL 703

RESULT 25

ENTRY 514210

GENETICS \*type complete

ORGANISM yopD protein - Yersinia enterocolitica

DATE 25-Feb-1994 sequence\_revision 10-Nov-1995 text\_change

ACCESSIONS 514210

REFERENCE 514210

author Michiels, T.; Wattiau, P.; Brasseur, R.; Buysschaert, J.M.

journal J. Gen. Microbiol. (1990) 58:2840-2849

title Secretion of Yop proteins by *Yersinia*.

cross-references DB:55655; NID:903104

697 LCA5VL 703

RESULT 23

ENTRY 504916

GENETICS \*type fragment

ORGANISM T-cell receptor delta chain precursor V-D-J region (clone

DATE 30-Sep-1991 sequence\_revision 30-Sep-1991 text\_change

ACCESSIONS 504916

REFERENCE 504916

author Morita, Y.; Reinmann, J.; Michalopoulos, E.; Ciccone, E.

journal J. Exp. Med. (1989) 169:393-405

title Diversity and structure of human T cell receptor delta chain

accession 504916

molecule\_type DNA

residues 1118

cross-references DBL:273147; NID:91560316; PID:91560317;

cross-references DBL:273147; NID:91560316; PID:91560317;

cross-references DBL:273147; NID:91560316; PID:91560317;

CLASSIFICATION

REMARKS

SUMMARY length 168 checksum 1221

Query Match 0.7% Score 7; DB 2; Length 168;

Best Local Similarity 100.0%; Pred. No. 1.04e+01;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 4 LCA5VL 10

697 LCA5VL 703

RESULT 26

ENTRY 514210

GENETICS \*type complete

ORGANISM myosin alkali light chain 3, ventricular and slow skeletal

DATE 12-Feb-1993 sequence\_revision 03-Oct-1995 text\_change

ACCESSIONS 509571; 515759; 506169

REFERENCE 509571; 515759; 506169

author Kuriyama, T.; Kuriyama, T.; Kuriyama, T.

journal J. Biol. Chem. (1993) 268:17373-17376

title Characterization of a rat myosin alkali light chain gene

accession 515759

molecule\_type DNA

residues 1118

cross-references DBL:273147; NID:91560316; PID:91560317;

cross-references DBL:273147; NID:91560316; PID:91560317;

cross-references DBL:273147; NID:91560316; PID:91560317;

GENETICS

gene myosin

REMARKS

SUMMARY length 182 molecule\_weight 20828 checksum 1492

Query Match 0.7% Score 7; DB 2; Length 182;

Best Local Similarity 100.0%; Pred. No. 1.04e+01;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 63 G51LTL 68

697 LCA5VL 703

RESULT 27

ENTRY 514210

GENETICS \*type complete

ORGANISM myosin alkali light chain 3, ventricular and slow skeletal

DATE 12-Feb-1993 sequence\_revision 03-Oct-1995 text\_change

ACCESSIONS 509571; 515759; 506169

REFERENCE 509571; 515759; 506169

author Kuriyama, T.; Kuriyama, T.; Kuriyama, T.

journal J. Biol. Chem. (1993) 268:17373-17376

title Characterization of a rat myosin alkali light chain gene

accession 515759

molecule\_type DNA

residues 1118

cross-references DBL:273147; NID:91560316; PID:91560317;

cross-references DBL:273147; NID:91560316; PID:91560317;

cross-references DBL:273147; NID:91560316; PID:91560317;

GENETICS

gene myosin

REMARKS

SUMMARY length 182 molecule\_weight 20828 checksum 1492

Query Match 0.7% Score 7; DB 2; Length 182;

Best Local Similarity 100.0%; Pred. No. 1.04e+01;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 63 G51LTL 68



RESULT 31  
ENTRY 554352  
TITLE alpha-tocopherol transfer protein - human  
ORGANISM Homo sapiens  
DATE 10-Oct-1997  
ACCESSIONS 554352  
REFERENCE Bloch, M. (1995) 306:437-443  
\*Journal Human alpha-tocopherol transfer protein: cDNA cloning, expression and chromosomal localization.  
\*Title Human alpha-tocopherol transfer protein: cDNA cloning.  
\*Accession 554352  
\*Status preliminary  
\*Molecule-type mRNA  
\*Residues 1-278  
\*Cross-references CDB:513364; OMIM:600415  
\*Gene CDB:TPP; AVE0  
\*Position Eq13.1-Eq13.3  
\*Location Superficially cellular retinaldehyde-binding protein homology  
\*Feature 53-747  
\*Label CDB  
\*Length 278  
\*Molecular-weight 31749  
\*Checksum 9071  
Query Match 0.74; Score 7; DB 2; Length 278;  
Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 36 NCAPVL 42  
07 230 NCAPVL 236

RESULT 32  
ENTRY 869337  
TITLE conserved hypothetical protein AF2298 - Archaeoglobus  
ORGANISM Archaeoglobus fulgidus  
DATE 05-Dec-1997  
ACCESSIONS 869337  
REFERENCE H. P. Clayton, R. A. Tomb, J. P. White, O. Nelson, K. E. Ketchum, K. A. Dodson, R. J. Gavett, M. Hickey, E. K. Peterson, J. D. Richardson, D. L. Karpavicz, A. R. Graham, D. E. Kyrle, R. C. Fleischmann, R. D. Quake, S. J. Lee, M. H. Sutton, C. O. Gill, S. J. Kirschner, E. V.  
\*Accession 869337  
\*Status preliminary  
\*Molecule-type DNA  
\*Residues 1-286  
\*Cross-references EMBL:U00003; MIM:945411  
\*Gene AF2298  
\*Position 1-286  
\*Location Superficially cellular retinaldehyde-binding protein homology  
\*Feature 53-747  
\*Label CDB  
\*Length 286  
\*Molecular-weight 32861  
\*Checksum 5503  
Query Match 0.74; Score 7; DB 2; Length 286;  
Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 36 NCAPVL 42  
07 230 NCAPVL 236

RESULT 33  
ENTRY 516881  
TITLE alpha-tocopherol transfer protein - human  
ORGANISM Homo sapiens  
DATE 10-Oct-1997  
ACCESSIONS 516881  
REFERENCE Bloch, M. (1995) 306:437-443  
\*Journal Human alpha-tocopherol transfer protein: cDNA cloning, expression and chromosomal localization.  
\*Title Human alpha-tocopherol transfer protein: cDNA cloning.  
\*Accession 516881  
\*Status preliminary  
\*Molecule-type mRNA  
\*Residues 1-278  
\*Cross-references CDB:513364; OMIM:600415  
\*Gene CDB:TPP; AVE0  
\*Position Eq13.1-Eq13.3  
\*Location Superficially cellular retinaldehyde-binding protein homology  
\*Feature 53-747  
\*Label CDB  
\*Length 278  
\*Molecular-weight 31749  
\*Checksum 9071  
Query Match 0.74; Score 7; DB 2; Length 278;  
Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 36 NCAPVL 42  
07 230 NCAPVL 236

RESULT 34  
ENTRY 516881  
TITLE alpha-tocopherol transfer protein - human  
ORGANISM Homo sapiens  
DATE 10-Oct-1997  
ACCESSIONS 516881  
REFERENCE Bloch, M. (1995) 306:437-443  
\*Journal Human alpha-tocopherol transfer protein: cDNA cloning, expression and chromosomal localization.  
\*Title Human alpha-tocopherol transfer protein: cDNA cloning.  
\*Accession 516881  
\*Status preliminary  
\*Molecule-type mRNA  
\*Residues 1-278  
\*Cross-references CDB:513364; OMIM:600415  
\*Gene CDB:TPP; AVE0  
\*Position Eq13.1-Eq13.3  
\*Location Superficially cellular retinaldehyde-binding protein homology  
\*Feature 53-747  
\*Label CDB  
\*Length 278  
\*Molecular-weight 31749  
\*Checksum 9071  
Query Match 0.74; Score 7; DB 2; Length 278;  
Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 36 NCAPVL 42  
07 230 NCAPVL 236

RESULT 35  
ENTRY 516881  
TITLE alpha-tocopherol transfer protein - human  
ORGANISM Homo sapiens  
DATE 10-Oct-1997  
ACCESSIONS 516881  
REFERENCE Bloch, M. (1995) 306:437-443  
\*Journal Human alpha-tocopherol transfer protein: cDNA cloning, expression and chromosomal localization.  
\*Title Human alpha-tocopherol transfer protein: cDNA cloning.  
\*Accession 516881  
\*Status preliminary  
\*Molecule-type mRNA  
\*Residues 1-278  
\*Cross-references CDB:513364; OMIM:600415  
\*Gene CDB:TPP; AVE0  
\*Position Eq13.1-Eq13.3  
\*Location Superficially cellular retinaldehyde-binding protein homology  
\*Feature 53-747  
\*Label CDB  
\*Length 278  
\*Molecular-weight 31749  
\*Checksum 9071  
Query Match 0.74; Score 7; DB 2; Length 278;  
Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 36 NCAPVL 42  
07 230 NCAPVL 236

RESULT 36  
ENTRY 516881  
TITLE alpha-tocopherol transfer protein - human  
ORGANISM Homo sapiens  
DATE 10-Oct-1997  
ACCESSIONS 516881  
REFERENCE Bloch, M. (1995) 306:437-443  
\*Journal Human alpha-tocopherol transfer protein: cDNA cloning, expression and chromosomal localization.  
\*Title Human alpha-tocopherol transfer protein: cDNA cloning.  
\*Accession 516881  
\*Status preliminary  
\*Molecule-type mRNA  
\*Residues 1-278  
\*Cross-references CDB:513364; OMIM:600415  
\*Gene CDB:TPP; AVE0  
\*Position Eq13.1-Eq13.3  
\*Location Superficially cellular retinaldehyde-binding protein homology  
\*Feature 53-747  
\*Label CDB  
\*Length 278  
\*Molecular-weight 31749  
\*Checksum 9071  
Query Match 0.74; Score 7; DB 2; Length 278;  
Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 36 NCAPVL 42  
07 230 NCAPVL 236

RESULT 37  
ENTRY 516881  
TITLE alpha-tocopherol transfer protein - human  
ORGANISM Homo sapiens  
DATE 10-Oct-1997  
ACCESSIONS 516881  
REFERENCE Bloch, M. (1995) 306:437-443  
\*Journal Human alpha-tocopherol transfer protein: cDNA cloning, expression and chromosomal localization.  
\*Title Human alpha-tocopherol transfer protein: cDNA cloning.  
\*Accession 516881  
\*Status preliminary  
\*Molecule-type mRNA  
\*Residues 1-278  
\*Cross-references CDB:513364; OMIM:600415  
\*Gene CDB:TPP; AVE0  
\*Position Eq13.1-Eq13.3  
\*Location Superficially cellular retinaldehyde-binding protein homology  
\*Feature 53-747  
\*Label CDB  
\*Length 278  
\*Molecular-weight 31749  
\*Checksum 9071  
Query Match 0.74; Score 7; DB 2; Length 278;  
Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 36 NCAPVL 42  
07 230 NCAPVL 236

RESULT 38  
ENTRY 516881  
TITLE alpha-tocopherol transfer protein - human  
ORGANISM Homo sapiens  
DATE 10-Oct-1997  
ACCESSIONS 516881  
REFERENCE Bloch, M. (1995) 306:437-443  
\*Journal Human alpha-tocopherol transfer protein: cDNA cloning, expression and chromosomal localization.  
\*Title Human alpha-tocopherol transfer protein: cDNA cloning.  
\*Accession 516881  
\*Status preliminary  
\*Molecule-type mRNA  
\*Residues 1-278  
\*Cross-references CDB:513364; OMIM:600415  
\*Gene CDB:TPP; AVE0  
\*Position Eq13.1-Eq13.3  
\*Location Superficially cellular retinaldehyde-binding protein homology  
\*Feature 53-747  
\*Label CDB  
\*Length 278  
\*Molecular-weight 31749  
\*Checksum 9071  
Query Match 0.74; Score 7; DB 2; Length 278;  
Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 36 NCAPVL 42  
07 230 NCAPVL 236

RESULT 39  
ENTRY 516881  
TITLE alpha-tocopherol transfer protein - human  
ORGANISM Homo sapiens  
DATE 10-Oct-1997  
ACCESSIONS 516881  
REFERENCE Bloch, M. (1995) 306:437-443  
\*Journal Human alpha-tocopherol transfer protein: cDNA cloning, expression and chromosomal localization.  
\*Title Human alpha-tocopherol transfer protein: cDNA cloning.  
\*Accession 516881  
\*Status preliminary  
\*Molecule-type mRNA  
\*Residues 1-278  
\*Cross-references CDB:513364; OMIM:600415  
\*Gene CDB:TPP; AVE0  
\*Position Eq13.1-Eq13.3  
\*Location Superficially cellular retinaldehyde-binding protein homology  
\*Feature 53-747  
\*Label CDB  
\*Length 278  
\*Molecular-weight 31749  
\*Checksum 9071  
Query Match 0.74; Score 7; DB 2; Length 278;  
Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 36 NCAPVL 42  
07 230 NCAPVL 236

RESULT 40  
ENTRY 516881  
TITLE alpha-tocopherol transfer protein - human  
ORGANISM Homo sapiens  
DATE 10-Oct-1997  
ACCESSIONS 516881  
REFERENCE Bloch, M. (1995) 306:437-443  
\*Journal Human alpha-tocopherol transfer protein: cDNA cloning, expression and chromosomal localization.  
\*Title Human alpha-tocopherol transfer protein: cDNA cloning.  
\*Accession 516881  
\*Status preliminary  
\*Molecule-type mRNA  
\*Residues 1-278  
\*Cross-references CDB:513364; OMIM:600415  
\*Gene CDB:TPP; AVE0  
\*Position Eq13.1-Eq13.3  
\*Location Superficially cellular retinaldehyde-binding protein homology  
\*Feature 53-747  
\*Label CDB  
\*Length 278  
\*Molecular-weight 31749  
\*Checksum 9071  
Query Match 0.74; Score 7; DB 2; Length 278;  
Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 36 NCAPVL 42  
07 230 NCAPVL 236

title The GMP-55 genome sequence of *Escherichia coli* K-12.  
 accession NC001742.607  
 status preliminary; nucleic acid sequence not shown;  
 molecule-type DNA  
 references DB:AF000277; DB:U00096; NID:91787543; PID:91787550;  
 experimental\_source strain K-12, substrain MO1555  
 GENETICS  
 gene smp  
 SUMPART length 321 molecular-weight 36038 checksum 1919  
 Query Match 0.7% Score 7; DB 2; Length 321  
 Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 83 LAF0PL 89  
 QY 119 LAF0PL 125  
 RESULT 37  
 ENTRY type complete  
 title smp protein - *Salmonella typhimurium*  
 ORIGINISM  
 formal\_name *Salmonella typhimurium*  
 20-Feb-1995 sequence\_revision 20-Feb-1995 text\_change  
 ACCESSIONS  
 REFERENCE  
 author Parra-Lopez, C.; Baer, M.T.; Grolman, R.A.  
 journal EMBO J. (1993) 12:4053-4062  
 title Molecular cloning of a locus required for resistance  
 to antitubercular peptides in *Salmonella typhimurium*.  
 accession 539586  
 molecule-type DNA 321 slabel P8  
 references DB:U00096; NID:914207; PID:914209  
 GENETICS  
 gene smp  
 SUMPART length 321 molecular-weight 36083 checksum 3845  
 Query Match 0.7% Score 7; DB 2; Length 321;  
 Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 83 LAF0PL 89  
 QY 11111111

ORIGINISM  
 formal\_name *Schistosoma mansoni* genome  
 31-Oct-1997  
 ACCESSIONS  
 REFERENCE  
 author S. S. Chiriac, C. M.  
 journal submitted to the EMBL Data Library, July 1995  
 accession 558158  
 status preliminary  
 molecule-type DNA 328 slabel G8  
 references DB:U00096; NID:91052783; PID:91052797  
 SUMPART length 329 molecular-weight 36687 checksum 5307  
 Query Match 0.7% Score 7; DB 2; Length 329;  
 Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 287 ASRSPL 293  
 QY 250 ASRSPL 256  
 RESULT 40  
 ENTRY type fragment  
 title hemocytin, a protein complicating haemochromatosis  
 ORIGINISM  
 formal\_name Homo sapiens (common name man)  
 17-Dec-1996 sequence\_revision 21-Jan-1997 text\_change  
 ACCESSIONS  
 REFERENCE  
 author Reder, V.; Fletcher, L.; Powell, L.  
 journal Biochem. Biophys. Res. Commun. (1996) 226:461-466  
 title Molecular cloning and characterization of a human  
 hemochromatosis gene.  
 accession PC4211  
 molecule-type RNA  
 references DB:U00096; NID:914207; PID:914209  
 SUMPART length 342 checksum 4470  
 Query Match 0.7% Score 7; DB 2; Length 342;  
 Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 68 LAF0PL 74

QY 119 LAF0PL 125  
 RESULT 38  
 ENTRY type complete  
 title conserved hypothetical protein MTH894 - *Methanobacterium*  
 ORIGINISM  
 formal\_name *Methanobacterium thermoautotrophicum* (strain Delta H)  
 05-Dec-1997 sequence\_revision 05-Dec-1997 text\_change  
 ACCESSIONS  
 REFERENCE  
 author Smith, D.R.; Doncette-Stamm, L.A.; Delouhery, C.; Lee, R.;  
 Dubois, J.; Blagden, T.; Mashiradeh, R.; Blakey, D. P.;  
 Lunn, W.; Pochter, B.; Qiu, D.; Spadofora, R.; Viscari, R.;  
 Wang, Y.; Wierzbowski, J.; Gibson, R.; Jivani, N.; Caruso, A.;  
 Bush, D.; Sater, R.; Patwell, D.; Pabst, S.;  
 Chong, G.; Sliemers, C.; Long, A.; Rice, P.; Hocking, J.;  
 Reeve, J.N. (1997) 179:7135-7155  
 J. Bacteriol. (1997) 179:7135-7155  
 title Complete genome sequence of *Methanobacterium*  
 thermoautotrophicum (strain Delta H): functional analysis and  
 comparative genomics.  
 accession E69219  
 molecule-type DNA translation not shown  
 status preliminary; nucleic acid sequence not shown;  
 references DB:U00096; NID:98037514  
 GENETICS  
 gene smp  
 SUMPART length 327 molecular-weight 35996 checksum 2247  
 Query Match 0.7% Score 7; DB 2; Length 327;  
 Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 101 ENCPYD 109  
 QY 231 ENCPYD 237  
 RESULT 39  
 ENTRY type complete  
 title hypothetical protein SPAC27.14c - *Clasium yeast*  
 ORIGINISM  
 formal\_name *Schistosoma mansoni* genome  
 31-Oct-1997

ORIGINISM  
 formal\_name *Schistosoma mansoni* genome  
 31-Oct-1997  
 ACCESSIONS  
 REFERENCE  
 author S. S. Chiriac, C. M.  
 journal submitted to the EMBL Data Library, July 1995  
 accession 558158  
 status preliminary  
 molecule-type DNA 328 slabel G8  
 references DB:U00096; NID:91052783; PID:91052797  
 SUMPART length 329 molecular-weight 36687 checksum 5307  
 Query Match 0.7% Score 7; DB 2; Length 329;  
 Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 287 ASRSPL 293  
 QY 250 ASRSPL 256  
 RESULT 41  
 ENTRY type complete  
 title hypothetical protein YJL023c - *Yeast (Saccharomyces cerevisiae)*  
 ORIGINISM  
 formal\_name *Saccharomyces cerevisiae*  
 06-Sep-1995  
 06-Sep-1995  
 accession 556793  
 status preliminary  
 molecule-type DNA  
 references DB:U00096; NID:91067144; PID:91067145; MIPS:YJL023c  
 SUMPART length 347 molecular-weight 39812 checksum 1050  
 Query Match 0.7% Score 7; DB 2; Length 347;  
 Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 67 STLDQ 73  
 QY 786 STLDQ 792  
 RESULT 42  
 ENTRY type complete  
 title UDP-glucose 4-epimerase (EC 5.1.3.2) - *Acetivillium brasiliense*  
 ORIGINISM  
 formal\_name *Acetivillium brasiliense*  
 19-Jul-1997  
 19-Jul-1997  
 accession 535409  
 status preliminary  
 molecule-type DNA  
 references DB:U00096; NID:914207; PID:914209  
 SUMPART length 342 checksum 4470  
 Query Match 0.7% Score 7; DB 2; Length 342;  
 Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 68 LAF0PL 74

```

****results****
1:348 #label: RES
****cross-references: EMDL:224076; RID:9396530; PID:9396532
COUNTRY: US
****gene****
CLASSIFICATION
REMARKS
PDBID: 11-315
SUMMARY
Query Match 0.74; Score 7; DB 2; Length 348;
Domain UDP-glucose 4-epimerase homology label cmp
#length 348 molecular-weight 36700 checksum 5193
Base Local Similarity 100.00; Pred. No. 1.04e+01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 342 CGASERS 348
QY 247 CGASERS 233

****results****
3:3
ENTRY 43
PROTEIN 138920 #type complete
TITLE Prostaglandin H2 receptor - human
AUTHOR Kozanietz, M.; Smith, W.L.; Blumberg, P.M.
ORGANISM Homo sapiens
DATE 06-Sep-1996
ACCESSIONS 138920
REFERENCES 138920 J.W.; Bally, T.J.; Pepperc, D.J.; Pierce, K.L.;
Kozanietz, M.; Smith, W.L.; Blumberg, P.M.;
Rogers, R.A.; Donello, J.E.; Falckham, C.E.; Kedzie,
H.M.; Woodard, D.F.; Gill, D.W.
JOURNAL Mol. Pharmacol. (1994) 66:213-220
TITLE Prostaglandin H2 receptor with
characteristic of the pharmacologically defined EP2
subtype.
****cross-references: MGI:94359403
****statistics****
138920 preliminary; translated from cD/EMBL/DDJ
****nucleic-acid****
138920 #type mRNA
****structures****
1:350 #label: RES
****cross-references: EMDL:469319; PID:4612650
SUMMARY
#length 350 molecular-weight 39675 checksum 150
Query Match 0.74; Score 7; DB 2; Length 350;
Base Local Similarity 100.00; Pred. No. 1.04e+01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 245 PDRNRG 251
QY 243 PDRNRG 247

```

[illegible][illegible][illegible][illegible]

ACCESIONS 565585 57559  
 REFERENCE 565585  
 \*author Beyer, S.; Dieler, J.; Pieperberg, W.  
 \*journal Mol. Gen. Genet. (1996) 256:75-78.  
 \*title The gene for the hydroxymethyltransferase of  
 (E7H 2734): new operon and evidence for pathway-specific  
 regulation by STR.  
 \*accession 565585 preliminary  
 \*molecule-type DNA  
 \*features  
 \*cross-reference EMBL: X89010; NID:980763; PID:980764  
 SUMMARY  
 status atc  
 length 428 \*molecular-weight 4563 \*checksum 6026  
 QUERY MATCH 0.78; Score 7; DB 2; Length 428;  
 Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 269 DRYGAM 275  
 1111111  
 659 DRYGAM 655  
 RESULT 49  
 55335 \*type complete  
 \*accession 55335  
 \*journal Mol. Gen. Genet. (1995) 247:639-645  
 \*title The gene for the hydroxymethyltransferase of an  
 endo-beta-1,5-glucanase gene from the mycoparasitic fungus  
 Trichoderma harzianum.  
 \*accession 55335  
 \*molecule-type preliminary  
 \*features  
 \*cross-reference EMBL: X89010; NID:980763; PID:980764  
 SUMMARY  
 status atc  
 length 430 \*molecular-weight 48165 \*checksum 6844  
 QUERY MATCH 0.78; Score 7; DB 2; Length 430;  
 Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 160 ANYDNA 166

QUERY 50  
 767 ANYDNA 773  
 1111111  
 RESULT 50  
 \*type complete  
 \*accession A69076  
 \*journal Nitrogenase molybdenum-nitrogen protein, NifN subunit.  
 \*title Methanobacterium thermoautotrophicum (strain Delta H)  
 \*accession A69076  
 \*molecule-type DNA  
 \*features  
 \*cross-reference EMBL: X89075; NID:980775; PID:980776  
 SUMMARY  
 status atc  
 length 448 \*molecular-weight 49096 \*checksum 4810  
 QUERY MATCH 0.78; Score 7; DB 2; Length 448;  
 Best Local Similarity 100.0%; Pred. No. 1.04e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 56 ANYDNA 64  
 817 ANYDNA 823  
 Search completed: Fri Dec 18 18:50:22 1998  
 Job time : 152 secs.



















```

CG  R18H OR MYC2184.34.
OS  HYPOCAERTEIN TUBERCULOSIS.
OS  HYPOCAERTEIN: FILAMENTOUS: ACTINOMYCETALS: HYPOCAERTEINACEAE.
NR  1.1
NR  SEQUENCE FROM N.A.
NR  STRAIN-1379V.
NA  MORRIS L., BARRETS D., BARRELL B.C., BALMAREAN M.A.:
NA  "THE BIOCHEMICAL JOURNAL" 1987; 243: 1-10.
CC  -1- FUNCTION: RIBOFLAVIN SYNTHESIS IS A STRUCTURAL ENZYME COMPLEX
CC  CATALYZING THE FORMATION OF RIBOFLAVIN FROM 5-AMINO-6-(1'-D)-
CC  RIBITYL-AMINO-2-(4,1H)-3,6-DIMETHYL-8-L-GLUTAMINE AND L-3,4-DIHYDRO-2-
CC  SUBSTITUTED-5,6-DIHYDRO-8-L-GLUTAMINE. THE BETA-
CC  SUBUNIT CATALYZES THE CONVERSION OF 5-AMINO-6-(1'-D)-RIBITYL-
CC  AMINO-2-(4,1H)-3,6-DIHYDRO-8-L-GLUTAMINE WITH L-3,4-DIHYDRO-2'-BUTANO-
CC  4-PHOSPHATE YIELDING 6,7-DIHYDRO-8-L-LYSINE.
CC  -1- CATALYTIC ACTIVITY: 2,6,7-DIHYDRO-8-(1-D-RIBITYL)GLUTAMINE +
CC  GLUTAMATE + H2O = 6,7-DIHYDRO-8-L-LYSINE + 2-HYDROXY-3-
CC  OXOPYRUVATE + PHOSPHORAN.
CC  -1- COFACTOR: FLAVOPROSTEN.
CC  -1- SUBUNIT: FINAL STEP OF RIBOFLAVIN SYNTHESIS.
CC  -1- STIMULANT: B265577.
CC  -1- STIMULANT: B265577.
DR  ENBL; 280108; 2265577.
DR  RIBOFLAVIN BIOSYNTHESIS; TRANSFERASE; FLAVOPROTEIN.
SQ  SUBSEQUENCE 156 AA; 1550L NM; 998B086 CMC32.
Oy  0.7%: Score 7; DB 1; Length 156;
Oy  Best Local Similarity 100.0%; Pexd. 4.47e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Db  324 ALDADH 34
Oy  111111
Oy  124 ALDADH 110

RESULT 16
ID  P47_S1HXH STANDARD: PRT: 155 AA.
AC  P47_S1HXH.1995 (REV. 33, CREATED)
DT  01-NOV-1995 (REV. 33, LAST SEQUENCE UPDATE)
DT  01-FEB-1996 (REV. 33, LAST ANNOTATION UPDATE)
DE  CHROMOPROTEIN 305 RIBOSOMAL PROTEIN S7.
DE  SPIROCYTIN MAXIMA.
OS  CHLOROPLAST.
OC  CHLOROPHYTES.
OC  EUKARYOTA, PLANTAE, PINCOPIPHYTES, CHLOROPHYTES (GREEN ALGAE),
OC  CONJUGATEDPHOTOTROPHS: EUGENIDIALES, EUGENIDIALES, EUGENIDIALES.
OC  CONJUGATEDPHOTOTROPHS: EUGENIDIALES, EUGENIDIALES, EUGENIDIALES.
NR  SEQUENCE FROM N.A.
NR  LBY K.A., MARHART J.R.:
NR  J. PROTOOL. 19:500-505(1993).

```

CC	-1. SIMILARITY: BELONGS TO THE 87P FAMILY OF RIBOSOMAL PROTEINS.
DR	EMBL: U07932; GI:10152; ...
PR	PROSITE: P08005; R08004; 1.
NR	PROSITE: P08005; R08004; 1.
SR	SEQUENCE 155 AA; 17769 MW; 6596348 CRR32;
SO	SEQUENCE 155 AA; 17769 MW; 6596348 CRR32;
Query Match	0.78; Score 7; DB 1; Length 155;
Matches	Match 1: 0; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Db	7711111111
243	AREGOS 249
RESULT 17	
AC	SEQUENCE STANDARD: PRG: 174 AA.
AD	CS1443
DT	01-NOV-1997 (REL. 35, CHEATED)
DY	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DZ	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE	CROSSOVER JUNCTION. PROTEIN ENCODING THE CDS (C 3, 1, 22, 4) (HOLLAND JUNCTION NICKLEASE RVC) (HOLLAND JUNCTION RESOLVASE RVC).
GN	RVC.
OS	PROKODONAS AERGIOSUS.
OC	PROKODONACEAE.
OR	PROKODONACEAE.
RA	PROKODONACEAE.
RB	PROKODONACEAE.
RC	PROKODONACEAE.
RD	PROKODONACEAE.
RE	PROKODONACEAE.
RF	PROKODONACEAE.
RG	PROKODONACEAE.
RH	PROKODONACEAE.
RI	PROKODONACEAE.
RJ	PROKODONACEAE.
RK	PROKODONACEAE.
RL	PROKODONACEAE.
RM	PROKODONACEAE.
RN	PROKODONACEAE.
RO	PROKODONACEAE.
RP	PROKODONACEAE.
RQ	PROKODONACEAE.
RR	PROKODONACEAE.
RS	PROKODONACEAE.
RT	PROKODONACEAE.
RU	PROKODONACEAE.
RV	PROKODONACEAE.
RW	PROKODONACEAE.
RX	PROKODONACEAE.
RY	PROKODONACEAE.
RZ	PROKODONACEAE.
RA	PROKODONACEAE.
RB	PROKODONACEAE.
RC	PROKODONACEAE.
RD	PROKODONACEAE.
RE	PROKODONACEAE.
RF	PROKODONACEAE.
RG	PROKODONACEAE.
RH	PROKODONACEAE.
RI	PROKODONACEAE.
RJ	PROKODONACEAE.
RK	PROKODONACEAE.
RL	PROKODONACEAE.
RM	PROKODONACEAE.
RO	PROKODONACEAE.
RP	PROKODONACEAE.
RQ	PROKODONACEAE.
RR	PROKODONACEAE.
RS	PROKODONACEAE.
RT	PROKODONACEAE.
RU	PROKODONACEAE.
RV	PROKODONACEAE.
RW	PROKODONACEAE.
RX	PROKODONACEAE.
RY	PROKODONACEAE.
RZ	PROKODONACEAE.
RA	PROKODONACEAE.
RB	PROKODONACEAE.
RC	PROKODONACEAE.
RD	PROKODONACEAE.
RE	PROKODONACEAE.
RF	PROKODONACEAE.
RG	PROKODONACEAE.
RH	PROKODONACEAE.
RI	PROKODONACEAE.
RJ	PROKODONACEAE.
RK	PROKODONACEAE.
RL	PROKODONACEAE.
RM	PROKODONACEAE.
RO	PROKODONACEAE.
RP	PROKODONACEAE.
RQ	PROKODONACEAE.
RR	PROKODONACEAE.
RS	PROKODONACEAE.
RT	PROKODONACEAE.
RU	PROKODONACEAE.
RV	PROKODONACEAE.
RW	PROKODONACEAE.
RX	PROKODONACEAE.
RY	PROKODONACEAE.
RZ	PROKODONACEAE.
RA	PROKODONACEAE.
RB	PROKODONACEAE.
RC	PROKODONACEAE.
RD	PROKODONACEAE.
RE	PROKODONACEAE.
RF	PROKODONACEAE.
RG	PROKODONACEAE.
RH	PROKODONACEAE.
RI	PROKODONACEAE.
RJ	PROKODONACEAE.
RK	PROKODONACEAE.
RL	PROKODONACEAE.
RM	PROKODONACEAE.
RO	PROKODONACEAE.
RP	PROKODONACEAE.
RQ	PROKODONACEAE.
RR	PROKODONACEAE.
RS	PROKODONACEAE.
RT	PROKODONACEAE.
RU	PROKODONACEAE.
RV	PROKODONACEAE.
RW	PROKODONACEAE.
RX	PROKODONACEAE.
RY	PROKODONACEAE.
RZ	PROKODONACEAE.
RA	PROKODONACEAE.
RB	PROKODONACEAE.
RC	PROKODONACEAE.
RD	PROKODONACEAE.
RE	PROKODONACEAE.
RF	PROKODONACEAE.
RG	PROKODONACEAE.
RH	PROKODONACEAE.
RI	PROKODONACEAE.
RJ	PROKODONACEAE.
RK	PROKODONACEAE.
RL	PROKODONACEAE.
RM	PROKODONACEAE.
RO	PROKODONACEAE.
RP	PROKODONACEAE.
RQ	PROKODONACEAE.
RR	PROKODONACEAE.
RS	PROKODONACEAE.
RT	PROKODONACEAE.
RU	PROKODONACEAE.
RV	PROKODONACEAE.
RW	PROKODONACEAE.
RX	PROKODONACEAE.
RY	PROKODONACEAE.
RZ	PROKODONACEAE.
RA	PROKODONACEAE.
RB	PROKODONACEAE.
RC	PROKODONACEAE.
RD	PROKODONACEAE.
RE	PROKODONACEAE.
RF	PROKODONACEAE.
RG	PROKODONACEAE.
RH	PROKODONACEAE.
RI	PROKODONACEAE.
RJ	PROKODONACEAE.
RK	PROKODONACEAE.
RL	PROKODONACEAE.
RM	PROKODONACEAE.
RO	PROKODONACEAE.
RP	PROKODONACEAE.
RQ	PROKODONACEAE.
RR	PROKODONACEAE.
RS	PROKODONACEAE.
RT	PROKODONACEAE.
RU	PROKODONACEAE.
RV	PROKODONACEAE.
RW	PROKODONACEAE.
RX	PROKODONACEAE.
RY	PROKODONACEAE.
RZ	PROKODONACEAE.
RA	PROKODONACEAE.
RB	PROKODONACEAE.
RC	PROKODONACEAE.
RD	PROKODONACEAE.
RE	PROKODONACEAE.
RF	PROKODONACEAE.
RG	PROKODONACEAE.
RH	PROKODONACEAE.
RI	PROKODONACEAE.
RJ	PROKODONACEAE.
RK	PROKODONACEAE.
RL	PROKODONACEAE.

Mon Dec 21 09:28:24 1991

US-08-951-733-14.rtf

Page 35

Mon Dec 21 09:28:24 1998

US-08-951-733-14.rtf

Page 40

[illegible][illegible]















**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**



















AC 015740: (REMBL. 05, CREATED)  
 DT 01-JAN-1998 (REMBL. 05, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (REMBL. 05, LAST ANNOTATION UPDATE)  
 DE PISA (FRAGMENT).  
 GN PISA (FRAGMENT).  
 OS PISCICULTURE DISORDER (LIVE WORLD).  
 OC EUTACTA PROCTON, SARCOMASTICOPHORA, SARODINA, RHODOPH.  
 RN (1).  
 RP SEQUENCE FROM N.A.  
 RA LOMIS M.F., JAMES N.;  
 RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL, AF019893, G2431115; .  
 FT NON\_TER 1  
 SQ SEQUENCE 680 AA: 76446 MW: 2828660 CRC32:  
 Query Match 0.84; Score 8; DB 5; Length 680;  
 Best Local Similarity 100.0%; Pred. No. 1.03e-01;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 166 REBLAF 173  
 QY 559 REBLAF 566  
 RESULT 17  
 ID 024144: (REMBL. 01, CREATED)  
 DT 01-NOV-1996 (REMBL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (REMBL. 01, LAST ANNOTATION UPDATE)  
 DE 17-NOV-1996 (REMBL. 05, LAST ANNOTATION UPDATE)  
 GN 17-NOV-1996 (REMBL. 05, LAST ANNOTATION UPDATE)  
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
 OC EUTACTA, METAZOA, ARTHROPODA, INSECTA, DIPTERA.  
 RN (1).  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE, 96194054.  
 RA PARK M.J., LUD J., SHARP E.J., ADLER P.H.;  
 RL DEVELOPMENT 127:951-969(1996).  
 DR EMBL, U71744, G1355007; .  
 FT NON\_TER 23  
 SQ SEQUENCE 869 AA: 97869 MW: 8395888 CRC32:  
 Query Match 0.84; Score 8; DB 5; Length 869;  
 Best Local Similarity 100.0%; Pred. No. 1.03e-01;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 430 SRSPLP 437

RP SEQUENCE FROM N.A.  
 RA MEDLINE, 96059477.  
 RA MADRUD L., GACIER M.C., VAYSSIE L., ROUANI A., SPERLING L.;  
 RL MOL. BIOL. CELL. 6:649-659(1995).  
 DR EMBL, U73510, G88110; .  
 FT NON\_TER 1  
 SQ SEQUENCE 23 AA: 2706 MW: A872074 CRC32:  
 Query Match 0.74; Score 7; DB 5; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 1.01e-01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 7 VERTUL 13  
 QY 785 VERTUL 791  
 RESULT 20  
 ID 065832: (REMBL. 07, CREATED)  
 DT 01-NOV-1996 (REMBL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (REMBL. 07, LAST ANNOTATION UPDATE)  
 DE 01-NOV-1996 (REMBL. 07, LAST ANNOTATION UPDATE)  
 GN LYCOPERSICON ESCULENTUM (TOMATO).  
 OS LYCOPERSICON ESCULENTUM (TOMATO).  
 OC EUTACTA, PLANTA, EMBRYOPHYTA, ANGIOSPERMAE, DIOCOTYLEDONEAE.  
 RN (1).  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-CV, AILSA CRAIG;  
 RA AGARIS A., KARELIS A.K.;  
 RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL, AF004926, G1184108; .  
 FT NON\_TER 1  
 SQ SEQUENCE 37 AA: 4405 MW: 5762180 CRC32:  
 Query Match 0.74; Score 7; DB 10; Length 37;  
 Best Local Similarity 100.0%; Pred. No. 1.01e-01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 5 RPSFLS 11  
 QY 364 RPSFLS 370  
 RESULT 21

QY 251 SRSPLP 258  
 RESULT 18  
 ID 060300: (REMBL. 07, CREATED)  
 DT 01-NOV-1996 (REMBL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (REMBL. 07, LAST ANNOTATION UPDATE)  
 DE 01-NOV-1996 (REMBL. 07, LAST ANNOTATION UPDATE)  
 GN KILAO553 PROTEIN (FRAGMENT).  
 OS KILAO553 PROTEIN (FRAGMENT).  
 OC EUTACTA, METAZOA, CHORDATA, VERTEBRATA, TETRAPODA, MAMMALIA.  
 RN (1).  
 RP SEQUENCE FROM N.A.  
 RA TISSUE-BRAIN.  
 RA MAGAZE T., ISHIVANA K., MIYAJIMA N., TANAKA A., KOTANI H., MONGHA N.;  
 RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL, AF019893, G2431115; .  
 FT NON\_TER 1  
 SQ SEQUENCE 1095 AA: 118835 MW: 43474254 CRC32:  
 Query Match 0.84; Score 8; DB 4; Length 1095;  
 Best Local Similarity 100.0%; Pred. No. 1.03e-01;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 941 ALAPTRA 948  
 QY 15 ALAPTRA 22  
 RESULT 19  
 ID 027174: (REMBL. 01, CREATED)  
 DT 01-NOV-1996 (REMBL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (REMBL. 01, LAST ANNOTATION UPDATE)  
 DE 01-NOV-1996 (REMBL. 05, LAST ANNOTATION UPDATE)  
 GN TRICHOCEST MATRIX PROTEIN T2 (FRAGMENT).  
 OS PARASCARID TETRAURELIA.  
 OC EUTACTA, PROTOSTOA, CILIOPHORA, CILIATA, HOLOTRICHIA, HYMENOSTOMATIDA.  
 RN (1).  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE, 95119139.  
 RA BARNES D.E., KARSISER S.P., LAMPEL K.A.;  
 RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL, AF025795, G268936; .  
 FT NON\_TER 87  
 SQ SEQUENCE 87 AA: 9613 MW: 3262580 CRC32:  
 Query Match 0.74; Score 7; DB 2; Length 87;  
 Best Local Similarity 100.0%; Pred. No. 1.01e-01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 52 LVAPOCA 58  
 QY 183 LVAPOCA 189  
 RESULT 22  
 ID 033890: (REMBL. 01, CREATED)  
 DT 01-NOV-1996 (REMBL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (REMBL. 01, LAST ANNOTATION UPDATE)  
 DE 01-NOV-1996 (REMBL. 01, LAST ANNOTATION UPDATE)  
 GN BMTA (FRAGMENT).  
 OS BMTA (FRAGMENT).  
 OC EUTACTA, PLANTA, EMBRYOPHYTA, ANGIOSPERMAE, DIOCOTYLEDONEAE.  
 RN (1).  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-CV, AILSA CRAIG;  
 RA AGARIS A., KARELIS A.K.;  
 RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL, AF004926, G1184108; .  
 FT NON\_TER 1  
 SQ SEQUENCE 109 AA: 11728 MW: 40093548 CRC32:  
 Query Match 0.74; Score 7; DB 3; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 1.01e-01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



[illegible]

Mon Dec 21 09:28:25 1998

US-08-951-733-14.rapt

Page 47

```

Query Match
Best Local Similarity 100.0%   Score 7:  DB 14: Length 149:
Matches 7:  Conservative      0:  Mismatches 0:  Indels 0:  Gaps 0

Db      86  SSISLDP 92
Qy      251  SSISLDP 257

RESULT 33
ID 083335      PRELIMINARY:  PRT:  149 MA.
DC 083335-1986 (TRIGLAMEL. 01, CREATED)
DT 01-NOV-1996 (TRIGLAMEL. 01, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TRIGLAMEL. 06, LAST ANNOTATION UPDATE)
T-CELL LYMPHOTOXIC VIRUS TYPE 2 REX AND TAX GENES, PARTIAL CDS, CLORE
DE SV2 (FLOANET).
GN REX.
GS REX.
OS HUMAN T-CELL LYMPHOTOXIC VIRUS TYPE II.
OC VIRAL: SS-RNA NONREPLICATED VIRUSES: POSITIVE STRAND RNA VIRUSES:
MAMMALIAN TYPE C ONCOVIRUS GROUP.
[1]
RP SEQUENCE FROM R.A.
RC REX.
EC MEDLINE 9610544.
RA STRAND N., NOVOT J., KORKEN C., OLIVERIA M.D., OLIVERIA O.D., ISHAK R.,
RA OLIVERIA M.P., LAMBEIRO P., ISHAK M., ACERVO V., BANERJEE S.M.,
RA ZHU D.M., ZHU D.S., ZHU D.L., WANG Y.
RA 2ND REX. 100% IDENTITY.
RA DB: 024873: 0163041: 7989.
FT NON_TER
SQ
SEQUENCE 149 MA: 15764 MW: 701046C CRC32:
Query Match
Best Local Similarity 100.0%   Score 7:  DB 14: Length 149:
Matches 7:  Conservative      0:  Mismatches 0:  Indels 0:  Gaps 0

Db      86  SSISLDP 92
Qy      251  SSISLDP 257

RESULT 34
ID 083333      PRELIMINARY:  PRT:  149 MA.
DC 083333-1986 (TRIGLAMEL. 01, CREATED)
DT 01-NOV-1996 (TRIGLAMEL. 01, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TRIGLAMEL. 06, LAST ANNOTATION UPDATE)
T-CELL LYMPHOTOXIC VIRUS TYPE 2 REX AND TAX GENES, PARTIAL CDS, CLORE
DE SV2 (FLOANET).

```

```

D2 PAR PX (FRAGMENT).
D3
D4
D5
D6
D7
D8
D9
D10
D11
D12
D13
D14
D15
D16
D17
D18
D19
D20
D21
D22
D23
D24
D25
D26
D27
D28
D29
D30
D31
D32
D33
D34
D35
D36
D37
D38
D39
D40
D41
D42
D43
D44
D45
D46
D47
D48
D49
D50
D51
D52
D53
D54
D55
D56
D57
D58
D59
D60
D61
D62
D63
D64
D65
D66
D67
D68
D69
D70
D71
D72
D73
D74
D75
D76
D77
D78
D79
D80
D81
D82
D83
D84
D85
D86
D87
D88
D89
D90
D91
D92
D93
D94
D95
D96
D97
D98
D99
D100
D101
D102
D103
D104
D105
D106
D107
D108
D109
D110
D111
D112
D113
D114
D115
D116
D117
D118
D119
D120
D121
D122
D123
D124
D125
D126
D127
D128
D129
D130
D131
D132
D133
D134
D135
D136
D137
D138
D139
D140
D141
D142
D143
D144
D145
D146
D147
D148
D149
D150
D151
D152
D153
D154
D155
D156
D157
D158
D159
D160
D161
D162
D163
D164
D165
D166
D167
D168
D169
D170
D171
D172
D173
D174
D175
D176
D177
D178
D179
D180
D181
D182
D183
D184
D185
D186
D187
D188
D189
D190
D191
D192
D193
D194
D195
D196
D197
D198
D199
D200
D201
D202
D203
D204
D205
D206
D207
D208
D209
D210
D211
D212
D213
D214
D215
D216
D217
D218
D219
D220
D221
D222
D223
D224
D225
D226
D227
D228
D229
D230
D231
D232
D233
D234
D235
D236
D237
D238
D239
D240
D241
D242
D243
D244
D245
D246
D247
D248
D249
D250
D251
D252
D253
D254
D255
D256
D257
D258
D259
D260
D261
D262
D263
D264
D265
D266
D267
D268
D269
D270
D271
D272
D273
D274
D275
D276
D277
D278
D279
D280
D281
D282
D283
D284
D285
D286
D287
D288
D289
D290
D291
D292
D293
D294
D295
D296
D297
D298
D299
D300
D301
D302
D303
D304
D305
D306
D307
D308
D309
D310
D311
D312
D313
D314
D315
D316
D317
D318
D319
D320
D321
D322
D323
D324
D325
D326
D327
D328
D329
D330
D331
D332
D333
D334
D335
D336
D337
D338
D339
D340
D341
D342
D343
D344
D345
D346
D347
D348
D349
D350
D351
D352
D353
D354
D355
D356
D357
D358
D359
D360
D361
D362
D363
D364
D365
D366
D367
D368
D369
D370
D371
D372
D373
D374
D375
D376
D377
D378
D379
D380
D381
D382
D383
D384
D385
D386
D387
D388
D389
D390
D391
D392
D393
D394
D395
D396
D397
D398
D399
D400
D401
D402
D403
D404
D405
D406
D407
D408
D409
D410
D411
D412
D413
D414
D415
D416
D417
D418
D419
D420
D421
D422
D423
D424
D425
D426
D427
D428
D429
D430
D431
D432
D433
D434
D435
D436
D437
D438
D439
D440
D441
D442
D443
D444
D445
D446
D447
D448
D449
D450
D451
D452
D453
D454
D455
D456
D457
D458
D459
D460
D461
D462
D463
D464
D465
D466
D467
D468
D469
D470
D471
D472
D473
D474
D475
D476
D477
D478
D479
D480
D481
D482
D483
D484
D485
D486
D487
D488
D489
D490
D491
D492
D493
D494
D495
D496
D497
D498
D499
D500
D501
D502
D503
D504
D505
D506
D507
D508
D509
D510
D511
D512
D513
D514
D515
D516
D517
D518
D519
D520
D521
D522
D523
D524
D525
D526
D527
D528
D529
D530
D531
D532
D533
D534
D535
D536
D537
D538
D539
D540
D541
D542
D543
D544
D545
D546
D547
D548
D549
D550
D551
D552
D553
D554
D555
D556
D557
D558
D559
D560
D561
D562
D563
D564
D565
D566
D567
D568
D569
D570
D571
D572
D573
D574
D575
D576
D577
D578
D579
D580
D581
D582
D583
D584
D585
D586
D587
D588
D589
D590
D591
D592
D593
D594
D595
D596
D597
D598
D599
D600
D601
D602
D603
D604
D605
D606
D607
D608
D609
D610
D611
D612
D613
D614
D615
D616
D617
D618
D619
D620
D621
D622
D623
D624
D625
D626
D627
D628
D629
D630
D631
D632
D633
D634
D635
D636
D637
D638
D639
D640
D641
D642
D643
D644
D645
D646
D647
D648
D649
D650
D651
D652
D653
D654
D655
D656
D657
D658
D659
D660
D661
D662
D663
D664
D665
D666
D667
D668
D669
D670
D671
D672
D673
D674
D675
D676
D677
D678
D679
D680
D681
D682
D683
D684
D685
D686
D687
D688
D689
D690
D691
D692
D693
D694
D695
D696
D697
D698
D699
D700
D701
D702
D703
D704
D705
D706
D707
D708
D709
D710
D711
D712
D713
D714
D715
D716
D717
D718
D719
D720
D721
D722
D723
D724
D725
D726
D727
D728
D729
D730
D731
D732
D733
D734
D735
D736
D737
D738
D739
D740
D741
D742
D743
D744
D745
D746
D747
D748
D749
D750
D751
D752
D753
D754
D755
D756
D757
D758
D759
D760
D761
D762
D763
D764
D765
D766
D767
D768
D769
D770
D771
D772
D773
D774
D775
D776
D777
D778
D779
D780
D781
D782
D783
D784
D785
D786
D787
D788
D789
D790
D791
D792
D793
D794
D795
D796
D797
D798
D799
D800
D801
D802
D803
D804
D805
D806
D807
D808
D809
D810
D811
D812
D813
D814
D815
D816
D817
D818
D819
D820
D821
D822
D823
D824
D825
D826
D827
D828
D829
D830
D831
D832
D833
D834
D835
D836
D837
D838
D839
D84
```

MON Dec 21 09:28:25 1998

US-08-951-733-14.rep

Page 48

[illegible]

```

DB      86 RESULDP  92
        1111111
CY      251 RESULDP 257

RESULT  36
ID      Q83399      PRELIMINARY:      PRT: 149 AA.
AC      Q83399:1996 (TREMBLABLE, 01, CREATED)
DT      01-NOV-1996 (TREMBLABLE, 01, LAST SEQUENCE UPDATE)
DE      T-CELL LYMPHOPROLIFIC VIRUS TYPE 2 REX AND TAX GENES, PARTIAL CDS, CLONE
        KAY401 1 PX (FRAGMENT).
OS      HUMAN T-CELL LYMPHOPROLIFIC VIRUS TYPE 11.
OC      VIRALDE; SS-RNA NONENVELOPED VIRUSES; POSITIVE STRAND RNA VIRUSES;
        RETROVIRAL; ONCOTRANSLA: TYPE C ONCOTRANS GROUP;
        [1]
        ONCOTRANSLA: TYPE C ONCOTRANS.
RM      SEQUENCE FROM R.A.
RX      READING: 5650504
RA      OLIVIERA, M.P., LAMBERTO, P., ISHAK M., ACEVEDO V., RAMIRESHAN, N.,
        KRU S.V., KIRO S.T., HALL, M.W.;
        J. VIROL. 70:1481-1492(1996).
RT      NON TERT
FT      50
        SEQUENCE 149 AA: 15779 MW: 58356CB CMC32;

Query March 07: Score 7; DB 14; Length 149;
        100.00% IDENTITY; 100.00% POSITIVE; 100.00% MATCHES;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0

DB      86 RESULDP  92
CY      251 RESULDP 257

RESULT  37
ID      Q39680      PRELIMINARY:      PRT: 153 AA.
AC      Q39680:
DT      01-NOV-1996 (TREMBLABLE, 01, CREATED)
DE      11-ND-1996 (TREMBLABLE, 01, LAST SEQUENCE UPDATE)
DE      HEAD-SHOCK COGNATE (FRAGMENT).
OS      DENDROCA. PLANTA. EXEMPTOPHYTA. ANTIOXIPERNAE: DICOTYLEDONALES: APYLALES;
        EUCALYPTUS.

```

```

AC 001198:
DT 01-JUL-1997 (TREMBALE, 04, CREATED)
DT 01-JUL-1998 (TREMBALE, 05, LAST SEQUENCE UPDATE)
DT 01-JUL-1998 (TREMBALE, 05, LAST ANNOTATION UPDATE)
DE STNAPROTEIN HOMOLOG.
OG HIRUDO MEDICINALIS (MEDICINAL LEECH).
OC EPHRAVOTA, METAZOA: ANNELIDA, HIRUDINIA.
NC GENUS FROM N.A.
NC TISSUE=NERVE COBD.
RX MEDLINE: 97197869
RX JOURNAL: ENCEPHALON 1998 110(1997).
DR J. HERCOWITZ: 019333501.
DR EMBL: U85805: G1933350.1
SO SEQUENCE 169 AA: 17847 MW: P4C68418 CRC32:

Query Match 0.74; Score 7; DB 5; Length 169;
Best Local Similarity 100.0%; Pred. No. 1.01e+01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0

DB 1A GLAPCA 20
GV 237 GLAPCA 243

RESULT 40
ID 073454 PRELIMINARY: PRT: 170 AA.
AC 073454:
DT 01-AUG-1998 (TREMBALE, 07, CREATED)
DT 01-AUG-1998 (TREMBALE, 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBALE, 07, LAST ANNOTATION UPDATE)
DE REX PROTEIN.
OG HUMAN T-CELL LYMPHOTROPIC VIRUS TYPE 2B
NC HUMAN T-CELL LYMPHOTROPIC VIRUS TYPE 2B
NC VIRUSES, RETROVIRUSES, RETROVIRINAE: BLV-HIV RETROVIRUSES,
OC HDNAV T-CELL LYMPHOTROPIC VIRUS TYPE 2.
[[1]]
NC GENUS FROM N.A.
NC TISSUE=NERVE COBD.
RX MEDLINE: 97197869
RX JOURNAL: ENCEPHALON 1998 110(1997).
DR J. HERCOWITZ: 019333501.
DR EMBL: U85805: G1933350.1
SO SEQUENCE 170 AA: 18042 MW: D645442 CRC32:

Query Match 0.74; Score 7; DB 14; Length 170;
Best Local Similarity 100.0%; Pred. No. 1.01e+01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0

```

```

DB      23 LEPIDON    29          0.78; Score 7; DB 10; Length 133;
Query Match                               Best Local Similarity 100.0%;
Matches       7; Conservative           0; Mismatches   0; Indels   0; Gaps   0;

OC      401 LEPIDON  407          PRELIMINARY:      PRT: 167 AA.

RESULT:  38
ID        NO.017729
AC Q07573;
DT 01-NOV-1996 (TRENDS&EEL, 01, CREATED)
DR 01-NOV-1996 (TRENDS&EEL, 01, LAST SEQUENCE UPDATE)
RS 01-NOV-1996 (TRENDS&EEL, 01, REVISIONS)
DE CHROMOSOME XII RADICAL FLAME ORF TL0422
CC SACHARONACEAE CEREVISIAE (BAKER'S YEAST).
OC ERIBATORIA, FUNGI; ASCOMYCOTINA, HEMIASCONECTES.
NM          SEQUENCE FROM N.A.
NA WEDLER R., WEDLER E., SCHARP N., VANBOTT R.;
RA SUBMITTED (MAY-1995) TO ENGL/GERBANH/DDBJ DATA BANKS.
RM          (1)
PA          (1)
SA MIPS;
SB          SEQUENCE FROM N.A.
SL ENBL: 273147; 22543970; --
SR ENBL: 273147; 22543970; --
SQ SEQUENCE 167 AA: 19759 HW: 5N08R574 CRC32:

Db      36 POLLANT    42          0.78; Score 7; DB 3; Length 167;
Query Match                               Best Local Similarity 100.0%;
Matches       0; Conservative           0; Mismatches   0; Indels   0; Gaps   0;

CY      882 POLLANT  888          PRELIMINARY:      PRT: 169 AA.

ID        NO.01386

```

```

Db      107 SHSLUP 113
Oy      251 SHSLUP 257

RESULT  41
AC Q08818 PRELIMINARY: PRT: 170 AA.
AD Q08818
DT 01-NOV-1996 (TRDBASEL 01, CREATED)
DT 01-NOV-1996 (TRDBASEL 02, LAST SEQUENCE UPDATE)
DT 01-SEP-1997 (TRDBASEL 01, LAST ANNOTATION UPDATE)
DT 01-SEP-1997 (TRDBASEL 02, LAST ANNOTATION UPDATE)
CM RNA
OS HUMAN T-CELL LEUKEMIA VIRUS TYPE II (ISOLATES G12 AND N9A) (HTLV-II).
OC VIRALDE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRINE.
RN (1)
RP SEQUENCE FROM N.A.
RK HEADLINE: 93323207.
RA N. A. POLDS T. A. SHILLER L. M. BADLOCK K. G., KAPLAN J. E., LAL R. B.,
RL J. VIROL. 67:4659-4664(1993).
RM (2)
RP SEQUENCE FROM N.A.
RA LEE R., IDLER K. B., SWANSON P., APARICIO J. J., CHIN K. K., LAM J. P.,
RA NOTER, H., MAN, T., LECKIE G., LANZETTA A., MAINICCI G., CHEN I. S. T.,
RA ROSINSKI, J. D.: 68(1993).
RD EMBL: L1456; G14840; -.
RD EMBL: L2073; G404041; -.
DS EMBL: L2073; G404041; -.
SO SEQUENCE 170 AA; 18973 MW; 662849EE CRC32;.

Oy      251 SHSLUP 257

Db      107 SHSLUP 113
Oy      251 SHSLUP 257

RESULT  42
AC Q08501 PRELIMINARY: PRT: 170 AA.
AD Q08501
DT 01-NOV-1996 (TRDBASEL 01, CREATED)
DT 01-NOV-1996 (TRDBASEL 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TRDBASEL 01, LAST ANNOTATION UPDATE)
DS REX 36 RD PROTEIN.
OS HUMAN T-CELL LYMPHOTROPIC VIRUS TYPE II.

```

```

OC VIRIDAE; SS-RNA NONENVELOPED VIRUSES; POSITIVE STRAND RNA VIRUSES;  
RETRONAVIRUS; ONCOTRANSFORMING TYPE C ONCOVIRUS GROUP;  
RNA [1]  
RN RETROVIRAL TYPE C ONCOVIRUSES.  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94250189.  
RY SUDANESSE R., PAMAKA R., PERINIS D., BRIGGS D., LEE T.R., ESSER M.,  
ZARULLO R.C., ROBERTSON W.N.A.;  
SCIENCE 225:421-424(1994).  
SR [2]  
SP SEQUENCE FROM N.A.  
SQ SEQUENCE FROM N.A.  
ST SILVERMORN E., TAKAHASHI Y., SHIMIZU N., GOTOBOHRI T., COULDE D.W.,  
CHEN T.S., MIYA H., SOUGIERE J.;  
PROC. NATL. ACAD. SCI. U.S.A. 82:3101-3105(1985).  
TA WEISS R.T., TEICH N., VAMOS R., COPPIN J.;  
J. VIROL. 67:1070-1083.  
TB (IN) WEISS R.T., TEICH N., VAMOS R., COPPIN J. (EDS.);  
RL RNA Tumor VIRUSES, SECOND EDITION, 2, VOL. 2:1070-1083;  
RM Cold Spring Harbor Laboratory, Cold Spring Harbor (1985).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA ROSENBLAT J.; (1985). NO DBOL/CGBANK/CDOS DATA BANKS.  
RB DBL: M1060: c129560:  
RC SEQUENCE 170 AA; 18463 MW; 00011932 CRC32;  
SD  
Seqy Match 0.7% Score 7; Rd 14; Length 170;  
Base Comp. Similarity 100.0%; Pctid No. 1,035+0;  
Matched 0; Conservative 0; Mismatches 0; Indels 0; Gaps 0  
Db 107 RESULT 113  
QY 251 RESULT 257
```

```

EX MEDLINE: 95286374.
RA HOLLISTON A. A., ROSOVIST R., WOLF-WANT H., POSSBERG A.,
RL 1995.
RP 1995.
RP 1995.
RA HOLLISTON A. A., 1994. TO DBL/GENBANK/DBS DATA BANKS.
RL DBL/1994/04, 0619941.
RA PLASMID.
SQ SEQUENCE 162 AA: 20959 MW: 47645FD CRC32.

Query Match 0.7% Score 7, Db 2, Length 162;
Best Local Similarity 100.0% Pctd. No. 1.01e+01;
Matches 7, Conservative 0, Mismatches 0, Indels 0, Gaps

Db 62 GILISTL 68
Oy 856 GILISTL 862

RESULT 44
ID 068704. PRELIMINARY: PRT, 162 AA.
DC 068704.188 (TRIMBLER, 07, CREDITED).
DC 01-NOV-1999 (TRIMBLER, 07, LAST SEQUENCE UPDATE)
D7 01-NOV-1999 (TRIMBLER, 07, LAST SEQUENCE UPDATE)
D7 01-NOV-1999 (TRIMBLER, 07, LAST SEQUENCE UPDATE)
DE TOP (TOPO) HOMOLOG (TOPO).
OS YERSINIA PESTIS.
OC PLASMID PCO1.
OC PNEUMONIA, CIRCULICTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS.
OC SCOTOBACTERIALE.
RN 11.
RP SEQUENCE FROM N. A.
NC STRAIN:NYU.
RA HOLLISTON A. A., WERDAY P. S., SEMENSKI P. J., GARRIS J.
RL 1995. TO DBL/GENBANK/DBS DATA BANKS.
RL SCHRITZER (MAR-1998) TO DBL/GENBANK/DBS DATA BANKS.
RA DBL/1995/09/06, 01956280.
RA PLASMID.
RV SEQUENCE 162 AA: 21000 MW: ADM431D7 CRC32.

Query Match 0.7% Score 7, Db 2, Length 162;
Best Local Similarity 100.0% Pctd. No. 1.01e+01;
Matches 7, Conservative 0, Mismatches 0, Indels 0, Gaps

Db 63 GILISTL 68
Oy 856 GILISTL 862

```

DB 70 TPLATLTY 76

[illegible]



**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

Release 3.1A John F. Collins, Bioinformatics Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

March:JP protein - protein database search, using Smith-Waterman algorithm

Run on: F71 Dec 18 19:04:41 1998; Nmapr time 37:18 Seconds  
Tabular output not generated.

Title: >US-08-951-733-20

Description: (1-1154) from OS08951733.pep

Sequence: 1 HNSGRCVLRHVALPVR.....TLMAAPALPSPFRITD 1154

Scoring table: TABLE uniprottable

Gap 60

Searched: 111922 seqs, 1610660 residues

Post-processing: Minimum Match 0%

Listing first 1000 summaries

Database: 4:geneseq2  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 3.325; Variance 0.726; scale 4.592

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Rank	Prod. No.	Score	Match	Length	ID	Description
48	6	0.5	17	25	W17555	Beta-2-cryotallin fr
49	6	0.5	20	28	W41112	Tax protein fragment
50	6	0.5	20	14	R12167	HTLV-1 p01(tax) prote
51	6	0.5	20	14	R12168	HTLV-1 p01(tax) prote
52	6	0.5	21	13	R73929	B-catenin-like CD epit
53	6	0.5	21	27	W51875	Leader sequence for u
54	6	0.5	21	29	R52639	Myotrophin tyrosine
55	6	0.5	22	19	R12167	HTLV-1 p01(tax) prote
56	6	0.5	22	19	R12168	HTLV-1 p01(tax) prote
57	6	0.5	26	19	W03637	C-protein-coupled odo
58	6	0.5	28	21	W11322	Generic human growth
59	6	0.5	31	20	W03990	Glucagon like peptide
60	6	0.5	31	23	W03990	Glucagon like peptide
61	6	0.5	31	23	W03990	Glucagon like peptide
62	6	0.5	31	23	W03990	Glucagon like peptide
63	6	0.5	31	23	W03990	Glucagon like peptide
64	6	0.5	31	23	W03990	Glucagon like peptide
65	6	0.5	31	23	W03990	Glucagon like peptide
66	6	0.5	31	23	W03990	Glucagon like peptide
67	6	0.5	31	23	W03990	Glucagon like peptide
68	6	0.5	31	23	W03990	Glucagon like peptide
69	6	0.5	31	23	W03990	Glucagon like peptide
70	6	0.5	31	23	W03990	Glucagon like peptide
71	6	0.5	31	23	W03990	Glucagon like peptide
72	6	0.5	31	23	W03990	Glucagon like peptide
73	6	0.5	31	23	W03990	Glucagon like peptide
74	6	0.5	31	23	W03990	Glucagon like peptide
75	6	0.5	31	23	W03990	Glucagon like peptide
76	6	0.5	31	23	W03990	Glucagon like peptide
77	6	0.5	31	23	W03990	Glucagon like peptide
78	6	0.5	31	23	W03990	Glucagon like peptide
79	6	0.5	31	23	W03990	Glucagon like peptide
80	6	0.5	31	23	W03990	Glucagon like peptide
81	6	0.5	31	23	W03990	Glucagon like peptide
82	6	0.5	31	23	W03990	Glucagon like peptide
83	6	0.5	31	23	W03990	Glucagon like peptide
84	6	0.5	31	23	W03990	Glucagon like peptide
85	6	0.5	31	23	W03990	Glucagon like peptide
86	6	0.5	31	23	W03990	Glucagon like peptide
87	6	0.5	31	23	W03990	Glucagon like peptide
88	6	0.5	31	23	W03990	Glucagon like peptide
89	6	0.5	31	23	W03990	Glucagon like peptide
90	6	0.5	31	23	W03990	Glucagon like peptide
91	6	0.5	31	23	W03990	Glucagon like peptide
92	6	0.5	31	23	W03990	Glucagon like peptide
93	6	0.5	31	23	W03990	Glucagon like peptide
94	6	0.5	31	23	W03990	Glucagon like peptide
95	6	0.5	31	23	W03990	Glucagon like peptide
96	6	0.5	31	23	W03990	Glucagon like peptide
97	6	0.5	31	23	W03990	Glucagon like peptide
98	6	0.5	31	23	W03990	Glucagon like peptide
99	6	0.5	31	23	W03990	Glucagon like peptide
100	6	0.5	31	23	W03990	Glucagon like peptide

Rank	Prod. No.	Score	Match	Length	ID	Description
1	7	0.6	126	9	R46882	CcBb cytochrome fused
2	7	0.6	136	20	W03991	Staphylokinase deriva
3	7	0.6	146	14	R41895	Sequence of human gta
4	7	0.6	146	14	R41895	Sequence of human gta
5	7	0.6	296	25	W03997	Human clone 55 protei
6	7	0.6	333	23	W19736	Sugar biosynthesis en
7	7	0.6	334	19	R45741	Myosin-like domain
8	7	0.6	334	19	R45741	Myosin-like domain
9	7	0.6	335	11	R08950	Recombinant high affi
10	7	0.6	355	6	R28972	Sequence in a high af
11	7	0.6	358	29	W44266	Human HPA proteoglyc
12	7	0.6	358	29	W44266	Human HPA proteoglyc
13	7	0.6	358	29	W44266	Human HPA proteoglyc
14	7	0.6	358	29	W44266	Human HPA proteoglyc
15	7	0.6	358	29	W44266	Human HPA proteoglyc
16	7	0.6	358	29	W44266	Human HPA proteoglyc
17	7	0.6	358	29	W44266	Human HPA proteoglyc
18	7	0.6	358	29	W44266	Human HPA proteoglyc
19	7	0.6	358	29	W44266	Human HPA proteoglyc
20	7	0.6	358	29	W44266	Human HPA proteoglyc
21	7	0.6	358	29	W44266	Human HPA proteoglyc
22	7	0.6	358	29	W44266	Human HPA proteoglyc
23	7	0.6	358	29	W44266	Human HPA proteoglyc
24	7	0.6	358	29	W44266	Human HPA proteoglyc
25	7	0.6	358	29	W44266	Human HPA proteoglyc
26	7	0.6	358	29	W44266	Human HPA proteoglyc
27	7	0.6	358	29	W44266	Human HPA proteoglyc
28	7	0.6	358	29	W44266	Human HPA proteoglyc
29	7	0.6	358	29	W44266	Human HPA proteoglyc
30	7	0.6	358	29	W44266	Human HPA proteoglyc
31	7	0.6	358	29	W44266	Human HPA proteoglyc
32	7	0.6	358	29	W44266	Human HPA proteoglyc
33	7	0.6	358	29	W44266	Human HPA proteoglyc
34	7	0.6	358	29	W44266	Human HPA proteoglyc
35	7	0.6	358	29	W44266	Human HPA proteoglyc
36	7	0.6	358	29	W44266	Human HPA proteoglyc
37	7	0.6	358	29	W44266	Human HPA proteoglyc
38	7	0.6	358	29	W44266	Human HPA proteoglyc
39	7	0.6	358	29	W44266	Human HPA proteoglyc
40	7	0.6	358	29	W44266	Human HPA proteoglyc
41	7	0.6	358	29	W44266	Human HPA proteoglyc
42	7	0.6	358	29	W44266	Human HPA proteoglyc
43	7	0.6	358	29	W44266	Human HPA proteoglyc
44	7	0.6	358	29	W44266	Human HPA proteoglyc
45	7	0.6	358	29	W44266	Human HPA proteoglyc
46	7	0.6	358	29	W44266	Human HPA proteoglyc
47	7	0.6	358	29	W44266	Human HPA proteoglyc











OS	Human lymphotropic virus type II strain RNA-19a.
PN	W0501457-21.
PR	10-JUN-1994; 007619.
PR	01-JUL-1993; 05-086413.
PR	20-JUN-1994; 05-259451.
PI	AB50-198077; Abam EM, Chen JCY, Edwards M, Colde DM, Gaidner P, Ialer M, Johnson AT, Lee HS, Mostley CT, Peterson B, Robertson E, Rosenblatt UJ, Stephens JE; Benson DA, 02/19/98.
PI	W5501457-21.
DR	N-5508; 081790.
P7	Human T-cell lymphotropic type II RNA viral genome, - expression products, and compositions for use in improved methods for detection
P3	Example 2C; Plasmid 100, 100bp; English.
CC	The sequence represents the rex region gene product (p26 protein) of the rex/tax region of the HTLV-II provirus RNA-19a strain genome.
CC	The rex region and tax region (encoding p10x protein, Rdg976) partially overlap and are transcribed from the same promoter.
CC	and mRNA processing. The virus may be produced in cell culture ATCC CRL 11360, and the encoded proteins may be produced as recombinant fusion proteins. The DNA and gene products may be used in diagnostic
CC	immunosorbent assays. HIV-100 antibodies can test samples by immunoprecipitation, Western blotting, and in antibody and vaccine
CC	preparation. 170 AA;
Q4	Sequence.
Query Match	0.63; Score 7; DB 16; Length 170;
Best Local Similarity	100.0%; Pval. 0.452e-01;
Matches	7; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Dh	107 seqalign 113
Qy	251 shrslup 257
RESULT	5
ID	W05397 standard; Protein: 295 AA.
DC	W05397-1988 (first entry)
DE	Human clone S5 protein
RV	src-homology region 3 domain; human; mouse; gsh domain; cell growth; cellular signaling element; cellular structural element; malignancy; protein identification; functional domain; protein screening;
OS	Human cellular transduction process.
OS	Romo sapiens.
PN	W0651625-31.
PR	10-OCT-1996.

[illegible]

```

57 gene -> organ production and engineering of myoinositol
58 dehydrogenase 8-9; 11bp; Japanese
59 CC The sequence (05700) shows a recombinant plasmid which encodes a
60 myoinositol dehydrogenase. The plasmid can be transformed into a
61 suitable host thus producing myoinositol dehydrogenase quickly and
62 easily.
63 Sequence 334 AA:
64
65 Query Match 0.63; Score 7; Db 9; Length 334;
66 Best Local Similarity 100.0%; Freq. 0.452e+01;
67 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
68
69 Db 98 atatega 104
70 11111111
71
72 304 ATSLDGA 310
73
74 R339.8
75 ID R339.8 standard; protein; 334 AA.
76 PF 21-FEB-1995 (first entry).
77 DT 21-FEB-1995 (first entry).
78 DZ Interleukin 8 (IL-8; receptor).
79 AC Interleukin; receptor; antiinflammatory; psoriasis; reperfusion;
80 rheumatoid arthritis; allograft; graft rejection.
81 KW Rheumatoid arthritis; allograft; graft rejection.
82 PR 065100395-8.
83
84 DR 12-APR-1994.
85 PF 21-MAY-1991; 1402455.
86 DT 21-MAY-1991 (first entry).
87 ID (UTRO) - UNIT ROTRON.
88 PA WPI; 94-1566657/19.
89 DR R-PSDB; 065919.
90
91 AC Interleukin-8 receptor polypeptide - having
92 P1 an extracellular domain, a transmembrane domain, and an intracellular
93 domain.
94 PS Class 1; Figure 1; 13bp; Japanese.
95 AC Anti-inflammatory agents containing the interleukin-8 receptor
96 polypeptide can be used for the treatment of psoriasis, rheumatoid
97 arthritis, and inflammatory diseases.
98 CC including reperfusion and allograft rejection.
99 Sequence 334 AA:
100
101 Query Match 0.63; Score 7; Db 11; Length 354;
102 Best Local Similarity 100.0%; Freq. 0.452e+01;
103 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
104
105 Db 335 lnhbrg 341
106 11111111
107
108 1104 LTRHRT 1110

```



50 Sequence 375 AA:  
Query Match 0.64; Score 7; DB 24; Length 375;  
Best Local Similarity 100.0%; Pred. No. 4,52e+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 260 cslp1d 266  
|||||  
QY 804 ETPSD 810

RESULT 14  
ID R80288 standard; Protein: 395 AA.  
AC R80288;  
DT 21-DEC-1995 (first entry)  
NM Galactose metabolism; enzyme: galK; gal operon; galactokinase.  
OS Streptomyces lividans.  
PN 25-0UL-1995;  
PR 25-0UL-1995; 834706  
PR 28-PED-1986; US-034706  
PR 30-JAN-1987; US-009419  
PR 30-JAN-1987; US-009419  
PR 30-JAN-1987; US-009419  
PR 30-JAN-1987; US-009419  
PR 30-JAN-1987; US-009419  
PA (ADAM) ADAMS C W  
PA (BRAN) BRANNER M E  
PA (FORM) FORMWALD J A  
PA (SCHM) SCHMIDT R J  
PA (WAL) WALSH J E  
DR N-PEMB; 098577.  
DR N-PEMB; 098577.  
PT New isolated Streptomyces coelicolor gal operon - used for the prodn. of  
PT recombinant DNA molecules for producing expression/cloning vectors in  
PT Example; Table 1; 2pp; English.  
CC A recombinant DNA molecule comprising a Streptomyces spp. gal operon  
CC promoter or any regulatable and functional deletion deriv. is  
CC promoter, galT, galK and galK structural genes and any other  
CC regulatory regions required for their transcription and translation.  
CC galT is uridine diphosphoglucose-4-epimerase; galK is  
CC galactokinase; and galS is galactose-1-phosphate uridylyltransferase.  
CC The gal operon is located on the 13.2 kb galactose operon of the  
CC region of Streptomyces lividans 1136. It was sequenced by chain  
CC termination and its complete DNA sequence is shown in 098577.  
CC The PI promoter is galactose inducible, glucose repressible and is  
CC the PI promoter for the entire operon. Data indicates  
CC that the operon product for the entire operon, galT, galK and  
CC galS. The transcript consists of approx. 1kb each for galT, galK and

50 Sequence 395 AA:  
Query Match 0.64; Score 7; DB 8; Length 395;  
Best Local Similarity 100.0%; Pred. No. 4,52e+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 375 PRGAP 381  
|||||  
QY 260 PRGAP 266

RESULT 15  
ID P70276 standard; Protein: 395 AA.  
AC P70276;  
DT 28-DEC-1992 (first entry)  
NM Galactose metabolism; promoter; expression vector; gal operon.  
OS Streptomyces lividans strain 1136.  
PN 07-SEP-1987;  
PR 07-SEP-1987; 870026  
PR 28-PED-1986; US-034706  
PR 30-JAN-1987; US-009419  
PR 30-JAN-1987; US-009419  
PA (SMK) SMITH K L  
PA (WAL) WALSH J E  
PA (BRAN) BRANNER M E  
PA (FORM) FORMWALD J A  
PA (SCHM) SCHMIDT R J  
PT New recombinant DNA contg. Streptomyces gal operon - or its  
PT promoter and gene fragments, and transformed cell able to express  
PT foreign protein and to metabolize galactose  
PT Description: Page 31-11, 170276; English known as galactose-1-phosphate  
CC uridylyltransferase (transferease), the gal gene product (P70276) is  
CC known as uridine diphosphoglucose-4-epimerase (epimerase), and the  
CC galK gene product (P70276) is known as galactose-1-kinase  
CC and the galS gene product (P70276) is known as galactose-1-phosphate  
CC uridylyltransferase (transferease). The gal operon is used for the  
CC expression of foreign DNA in a transformed micro-organism  
CC foreign DNA in a transformed micro-organism  
SQ Sequence 395 AA:

Query Match 0.64; Score 7; DB 3; Length 395;  
Best Local Similarity 100.0%; Pred. No. 4,52e+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 375 PRGAP 381  
|||||  
QY 260 PRGAP 266

RESULT 17  
ID W44252 standard; Protein: 398 AA.

CC galK. Galactose induction of PI is mediated, at least in part, by  
CC transcription start site and a repressor protein which recognizes  
CC the operator. The P2 promoter is upstream of the galT gene and  
CC transcribes both galT and galK genes. P2 promoter expression is  
CC constitutive.  
SQ Sequence 395 AA:

Query Match 0.64; Score 7; DB 14; Length 395;  
Best Local Similarity 100.0%; Pred. No. 4,52e+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 375 PRGAP 381  
|||||  
QY 260 PRGAP 266

RESULT 15  
ID R41530 standard; Protein: 395 AA.  
AC R41530;  
DT 28-DEC-1995 (first entry)  
NM Galactose metabolism; enzyme: galK; gal operon; galactokinase.  
OS Streptomyces lividans.  
PN 25-0UL-1995;  
PR 25-0UL-1995; 834706  
PR 28-PED-1986; US-034706  
PR 30-JAN-1987; US-009419  
PR 30-JAN-1987; US-009419  
PR 30-JAN-1987; US-009419  
PR 30-JAN-1987; US-009419  
PA (ADAM) ADAMS C W  
PA (BRAN) BRANNER M E  
PA (FORM) FORMWALD J A  
PA (SCHM) SCHMIDT R J  
PA (WAL) WALSH J E  
DR N-PEMB; 098577.  
DR N-PEMB; 098577.  
PT New isolated Streptomyces coelicolor gal operon sequences - including PI and P2  
PT promoters, galT, galS and galK genes, for enabling transformed  
PT hosts to utilize galactose  
PT Claim 2; Page 27-28; 3pp; English.  
CC The gal operon is located on the 13.2 kb galactose operon of the  
CC region of Streptomyces lividans 1136. It was sequenced by chain  
CC termination and its complete DNA sequence is shown in 098577.  
CC The PI promoter is galactose inducible, glucose repressible and is  
CC the PI promoter for the entire operon. Data indicates  
CC that the operon product for the entire operon, galT, galK and  
CC galS. The transcript consists of approx. 1kb each for galT, galK and  
CC galS. The transcript consists of approx. 1kb each for galT, galK and  
CC See also R41528 and R41529.

AC W44252;  
DT 18-DEC-1998 (first entry)  
NM Galactose metabolism; enzyme: galK; gal operon; galactokinase.  
OS Streptomyces lividans.  
PN 25-0UL-1995;  
PR 25-0UL-1995; 834706  
PR 28-PED-1986; US-034706  
PR 30-JAN-1987; US-009419  
PR 30-JAN-1987; US-009419  
PR 30-JAN-1987; US-009419  
PA (ADAM) ADAMS C W  
PA (BRAN) BRANNER M E  
PA (FORM) FORMWALD J A  
PA (SCHM) SCHMIDT R J  
PA (WAL) WALSH J E  
DR N-PEMB; 098577.  
DR N-PEMB; 098577.  
PT New recombinant DNA contg. Streptomyces gal operon - or its  
PT promoter and gene fragments, and transformed cell able to express  
PT foreign protein and to metabolize galactose  
PT Description: Page 31-11, 170276; English known as galactose-1-phosphate  
CC uridylyltransferase (transferease), the gal gene product (P70276) is  
CC known as uridine diphosphoglucose-4-epimerase (epimerase), and the  
CC galK gene product (P70276) is known as galactose-1-kinase  
CC and the galS gene product (P70276) is known as galactose-1-phosphate  
CC uridylyltransferase (transferease). The gal operon is used for the  
CC expression of foreign DNA in a transformed micro-organism  
CC foreign DNA in a transformed micro-organism  
SQ Sequence 398 AA:

Query Match 0.64; Score 7; DB 26; Length 398;  
Best Local Similarity 100.0%; Pred. No. 4,52e+01;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 157 LRAVTP 163  
|||||  
QY 437 LRAVTP 443

RESULT 18  
ID R88407 standard; Protein: 429 AA.  
AC R88407;  
DT 05-DEC-1996 (first entry)  
NM Galactose metabolism; enzyme: galK; gal operon; galactokinase.  
OS Streptomyces lividans.  
PN 25-0UL-1995;  
PR 25-0UL-1995; 834706  
PR 28-PED-1986; US-034706  
PR 30-JAN-1987; US-009419  
PR 30-JAN-1987; US-009419  
PR 30-JAN-1987; US-009419  
PA (ADAM) ADAMS C W  
PA (BRAN) BRANNER M E  
PA (FORM) FORMWALD J A  
PA (SCHM) SCHMIDT R J  
PA (WAL) WALSH J E  
DR N-PEMB; 098577.  
DR N-PEMB; 098577.  
PT New isolated Streptomyces coelicolor gal operon sequences - including PI and P2  
PT promoters, galT, galS and galK genes, for enabling transformed  
PT hosts to utilize galactose  
PT Claim 2; Page 27-28; 3pp; English.  
CC The gal operon is located on the 13.2 kb galactose operon of the  
CC region of Streptomyces lividans 1136. It was sequenced by chain  
CC termination and its complete DNA sequence is shown in 098577.  
CC The PI promoter is galactose inducible, glucose repressible and is  
CC the PI promoter for the entire operon. Data indicates  
CC that the operon product for the entire operon, galT, galK and  
CC galS. The transcript consists of approx. 1kb each for galT, galK and  
CC galS. The transcript consists of approx. 1kb each for galT, galK and  
CC See also R41528 and R41529.

Mon Dec 21 09:28:26 1998

Page 39

```

DR      N-PM28 090206.
DR      Production of very long chain fatty acids(1) in plant(s) - to produce
DR      drought and stress resistant transgenic plants(s)
PT      P
PT      CC Purification of jojoba wax synthase protein is described in PCT
PT      CC application WO 93/10241. Jojoba fatty acyl reductase sequence is
CC      known. A membrane prep. having wax synthase activity is isolated,
CC      the wax synthase activity is solubilized, and the wax synthase
CC      and seed coats of jojoba embryos.
CC      Sequence 493 M.
SQ
Query Match 0.6% Score 7; Db 14; Length 493;
      Best Local Similarity 100.0%; Pfad No.4,528+0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps
DB      426 111942 432
QY      884 LITERUD 890

RESULT 30
ID      J56470 standard; Protein: 493 M.
AC      R56470:
DC      14-NOV-1995 (first entry)
DE      Jojoba fatty acyl-CoA reductase.
DE      CC
DE      CC Simmondsia chinensis.
P1      085370396-A.
PD      05-DEC-1994
PF      21-FEB-1991, 659975
PF      21-FEB-1991, 659975
PR      21-FEB-1991; US-7962523.
PR      20-NOV-1991; US-7962526.
PR      31-JUL-1993; US-920430.
PA      (CAGL) CAGLGENE INC.
PA      P
DR      WPI: 95-021884/OJ 06; Pollard MR:
DR      N-PM28; Q73816.
DR      New recombinant constructs for transforming plants of E. coli
PT      P
PT      CC containing a nucleic acid encoding jojoba embryo long chain fatty
PT      CC chain 20. Figure 1: 30pp; English.
CC      The sequence encoding the jojoba embryo long chain fatty acyl-CoA
CC      reductase may be used in recombinant constructs which in turn can be
CC      expression of the jojoba embryo long chain fatty acyl-CoA reductase
CC      in host cells. The enzyme catalyses the formation of a fatty
CC      alcohol from a fatty acyl substrate.
CC      Sequence 493 M.

```

US-08-951-733-20.rag

Page 40

```

NM   PAR: PCR primer.
OS   Jojoba embryo.
LN   M921418.6.A
PR   21-DEC-1992; U01364.
PR   22-SEP-1991; US-659975.
PR   27-SEP-1991; US-767251.
PR   20-NOV-1991; US-782266.
PR   16-NOV-1991; US-782266.
PI   Metc JG, Pollard KM;
DR   WPI: 92-318370/38.
DR   N-PDB: Q24369. brassicic acid reductase, produced
PI   by recombinant DNA methods, actively convert fatty acyl
PI   substrates to fatty alcohols(s)
PS   Disclausier, F; 1(1-7): 86pp; English.
PS   A Jojoba cDNA library was screened using PCR primers
CC   designed for the Jojoba cDNA library. The cDNA for the
CC   reductase protein was obtained and sequenced. The
CC   sequence may be used in the produ. of PAR in recombinant
SQ   Sequence 493 AA;
Query Match      0.66; Score 7; DB 5; Length 493;
Best Local Statist 100.00; Pred. No. 4,52e+01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps
DB 426 1111111 432
QY 884 LITRBYD 890

RESULT 23
ID R79931 standard: Protein: 493 AA.
AC R79931;
PF 16-NOV-1991; 493aa (fatty acyl reductase)
PF 16-NOV-1991; 493aa (fatty acyl reductase)
PF Fatty acyl reductase; wax synthase; jojoba; oilseed;
PF Long-chain alcohol; fatty acyl; transesterase; wax ester;
PF transgenic plant; crop improvement; Brassica.
PF US5445947.7A chimera2.
PF 29-NOV-1995.
PF 20-NOV-1991; 786256.
PF 21-NOV-1992; US-831415.
PF 11-NOV-1992; WO-005863.
PF 20-MAY-1993; US-066299.
PI (CAL) J. Calkins;
PI (CAL) J. Calkins;
PI (CAL) J. Calkins;
PI WPI: 95-310894/40.
DR N-PDB: 7O4335.
DR

```













# WIRE

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd.

Wparch\_dp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Dec 18 19 01:39 1998; Wparch time 31.48 seconds  
Tabular output not generated.

Title: >US-08-951-731-20  
Description: 1.11154 from US08951731.pep  
Accession: 1 HAE0C9VLTWPAALPAT.....TLEMANPALPDEFTIIL 1154  
Sequence:

Scoring table: TABLE uninterpretable  
Gap 60

Searched: 107076 seqs, 3414558 residues

Post-processing: Minimum Match ON  
Detecting filter 1000 summaries

Database: p1556  
1: p151 2: p152 3: p153 4: p154

Statistics: Mean 4.071; Variance 0.451; scale 9.034

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	9	0.8	24	2	transmembrane glycopr	1.20e-03

2	9	0.8	183	2	A28135	1.30e-01
3	9	0.7	123	2	A28135	1.30e-01
4	8	0.7	169	2	A28135	1.30e-01
5	8	0.7	305	1	DG8529	1.44e-01
6	8	0.7	342	1	TMVW9	1.44e-01
7	8	0.7	400	1	TMVW9	1.44e-01
8	8	0.7	400	1	TMVW9	1.44e-01
9	8	0.7	508	2	CE2200	1.44e-01
10	8	0.7	654	1	H8Y81	1.44e-01
11	8	0.7	797	2	S28103	1.44e-01
12	8	0.7	797	2	S28103	1.44e-01
13	8	0.7	900	2	S28103	1.44e-01
14	8	0.7	900	2	S28103	1.44e-01
15	7	0.6	39	2	J00282	1.32e-01
16	7	0.6	52	2	A49196	1.32e-01
17	7	0.6	112	2	A49196	1.32e-01
18	7	0.6	112	2	A49196	1.32e-01
19	7	0.6	112	2	A49196	1.32e-01
20	7	0.6	115	2	A23339	1.32e-01
21	7	0.6	118	2	A23339	1.32e-01
22	7	0.6	148	2	S30260	1.32e-01
23	7	0.6	148	2	S30260	1.32e-01
24	7	0.6	167	2	S64794	1.32e-01
25	7	0.6	168	2	S64794	1.32e-01
26	7	0.6	168	2	S64794	1.32e-01
27	7	0.6	168	2	S64794	1.32e-01
28	7	0.6	168	2	S64794	1.32e-01
29	7	0.6	200	1	MORV9	1.32e-01
30	7	0.6	216	2	S73913	1.32e-01
31	7	0.6	216	2	S73913	1.32e-01
32	7	0.6	216	2	S73913	1.32e-01
33	7	0.6	216	2	S73913	1.32e-01
34	7	0.6	216	2	S73913	1.32e-01
35	7	0.6	216	2	S73913	1.32e-01
36	7	0.6	216	2	S73913	1.32e-01
37	7	0.6	216	2	S73913	1.32e-01
38	7	0.6	216	2	S73913	1.32e-01
39	7	0.6	216	2	S73913	1.32e-01
40	7	0.6	216	2	S73913	1.32e-01
41	7	0.6	216	2	S73913	1.32e-01
42	7	0.6	216	2	S73913	1.32e-01
43	7	0.6	216	2	S73913	1.32e-01
44	7	0.6	216	2	S73913	1.32e-01
45	7	0.6	216	2	S73913	1.32e-01
46	7	0.6	216	2	S73913	1.32e-01
47	7	0.6	216	2	S73913	1.32e-01
48	7	0.6	216	2	S73913	1.32e-01
49	7	0.6	216	2	S73913	1.32e-01
50	7	0.6	216	2	S73913	1.32e-01
51	7	0.6	216	2	S73913	1.32e-01
52	7	0.6	216	2	S73913	1.32e-01

53	7	0.6	385	2	S51087	1.32e-01
54	7	0.6	385	2	S51087	1.32e-01
55	7	0.6	385	2	S51087	1.32e-01
56	7	0.6	385	2	S51087	1.32e-01
57	7	0.6	412	2	U00438	1.32e-01
58	7	0.6	412	2	U00438	1.32e-01
59	7	0.6	412	2	U00438	1.32e-01
60	7	0.6	412	2	U00438	1.32e-01
61	7	0.6	412	2	U00438	1.32e-01
62	7	0.6	412	2	U00438	1.32e-01
63	7	0.6	412	2	U00438	1.32e-01
64	7	0.6	412	2	U00438	1.32e-01
65	7	0.6	412	2	U00438	1.32e-01
66	7	0.6	412	2	U00438	1.32e-01
67	7	0.6	412	2	U00438	1.32e-01
68	7	0.6	412	2	U00438	1.32e-01
69	7	0.6	412	2	U00438	1.32e-01
70	7	0.6	412	2	U00438	1.32e-01
71	7	0.6	412	2	U00438	1.32e-01
72	7	0.6	412	2	U00438	1.32e-01
73	7	0.6	412	2	U00438	1.32e-01
74	7	0.6	412	2	U00438	1.32e-01
75	7	0.6	412	2	U00438	1.32e-01
76	7	0.6	412	2	U00438	1.32e-01
77	7	0.6	412	2	U00438	1.32e-01
78	7	0.6	412	2	U00438	1.32e-01
79	7	0.6	412	2	U00438	1.32e-01
80	7	0.6	412	2	U00438	1.32e-01
81	7	0.6	412	2	U00438	1.32e-01
82	7	0.6	412	2	U00438	1.32e-01
83	7	0.6	412	2	U00438	1.32e-01
84	7	0.6	412	2	U00438	1.32e-01
85	7	0.6	412	2	U00438	1.32e-01
86	7	0.6	412	2	U00438	1.32e-01
87	7	0.6	412	2	U00438	1.32e-01
88	7	0.6	412	2	U00438	1.32e-01
89	7	0.6	412	2	U00438	1.32e-01
90	7	0.6	412	2	U00438	1.32e-01
91	7	0.6	412	2	U00438	1.32e-01
92	7	0.6	412	2	U00438	1.32e-01
93	7	0.6	412	2	U00438	1.32e-01
94	7	0.6	412	2	U00438	1.32e-01
95	7	0.6	412	2	U00438	1.32e-01
96	7	0.6	412	2	U00438	1.32e-01
97	7	0.6	412	2	U00438	1.32e-01
98	7	0.6	412	2	U00438	1.32e-01
99	7	0.6	412	2	U00438	1.32e-01
100	7	0.6	412	2	U00438	1.32e-01
101	7	0.6	412	2	U00438	1.32e-01
102	7	0.6	412	2	U00438	1.32e-01
103	7	0.6	412	2	U00438	1.32e-01

104	7	0.6	1366	1	CGH025	1.32e-01
105	7	0.6	1366	1	CGH025	1.32e-01
106	7	0.6	1366	1	CGH025	1.32e-01
107	7	0.6	1366	1	CGH025	1.32e-01
108	7	0.6	1366	1	CGH025	1.32e-01
109	7	0.6	1366	1	CGH025	1.32e-01
110	7	0.6	1366	1	CGH025	1.32e-01
111	7	0.6	1366	1	CGH025	1.32e-01
112	7	0.6	1366	1	CGH025	1.32e-01
113	7	0.6	1366	1	CGH025	1.32e-01
114	7	0.6	1366	1	CGH025	1.32e-01
115	7	0.6	1366	1	CGH025	1.32e-01
116	7	0.6	1366	1	CGH025	1.32e-01
117	7	0.6	1366	1	CGH025	1.32e-01
118	7	0.6	1366	1	CGH025	1.32e-01
119	7	0.6	1366	1	CGH025	1.32e-01
120	7	0.6	1366	1	CGH025	1.32e-01
121	7	0.6	1366	1	CGH025	1.32e-01
122	7	0.6	1366	1	CGH025	1.32e-01
123	7	0.6	1366	1	CGH025	1.32e-01
124	7	0.6	1366	1	CGH025	1.32e-01
125	7	0.6	1366	1	CGH025	1.32e-01
126	7	0.6	1366	1	CGH025	1.32e-01
127	7	0.6	1366	1	CGH025	1.32e-01
128	7	0.6	1366	1	CGH025	1.32e-01
129	7	0.6	1366	1	CGH025	1.32e-01
130	7	0.6	1366	1	CGH025	1.32e-01
131	7	0.6	1366	1	CGH025	1.32e-01
132	7	0.6	1366	1	CGH025	1.32e-01
133	7	0.6	1366	1	CGH025	1.32e-01
134	7	0.6	1366	1	CGH025	1.32e-01
135	7	0.6	1366	1	CGH025	1.32e-01
136	7	0.6	1366	1	CGH025	1.32e-01
137	7	0.6	1366	1	CGH025	1.32e-01
138	7	0.6	1366	1	CGH025	1.32e-01
139	7	0.6	1366	1	CGH025	1.32e-01
140	7	0.6	1366	1	CGH025	1.32e-01
141	7	0.6	1366	1	CGH025	1.32e-01
142	7	0.6	1366	1	CGH025	1.32e-01
143	7	0.6	1366	1	CGH025	1.32e-01
144	7	0.6	1366	1	CGH025	1.32e-01
145	7	0.6	1366	1	CGH025	1.32e-01
146	7	0.6	1366	1	CGH025	1.32e-01
147	7	0.6	1366	1	CGH025	1.32e-01
148	7	0.6	1366	1	CGH025	1.32e-01
149	7	0.6	1366	1	CGH025	1.32e-01
150	7	0.6	1366	1	CGH025	1.32e-01
151	7	0.6	1366	1	CGH025	1.32e-01
152	7	0.6	1366	1	CGH025	1.32e-01
153	7	0.6	1366	1	CGH025	1.32e-01
154	7	0.6	1366	1	CGH025	1.32e-01

155	6	0.5	125	2	150498	urotelin II-alpha pr	5.47e+02
156	6	0.5	125	2	150499	urotelin II-beta pr	5.47e+02
157	6	0.5	125	2	150500	urotelin II-gamma pr	5.47e+02
158	6	0.5	126	2	150488	urotelin II-delta pr	5.47e+02
159	6	0.5	130	2	573368	urotelin II-epsilon pr	5.47e+02
160	6	0.5	131	1	573369	urotelin II-zeta pr	5.47e+02
161	6	0.5	131	1	573370	urotelin II-eta pr	5.47e+02
162	6	0.5	132	2	159566	urotelin II-theta pr	5.47e+02
163	6	0.5	132	2	159567	urotelin II-iota pr	5.47e+02
164	6	0.5	133	2	159568	urotelin II-kappa pr	5.47e+02
165	6	0.5	134	2	159569	urotelin II-lambda pr	5.47e+02
166	6	0.5	135	2	159570	urotelin II-mu pr	5.47e+02
167	6	0.5	136	2	159571	urotelin II-nu pr	5.47e+02
168	6	0.5	137	2	159572	urotelin II-xi pr	5.47e+02
169	6	0.5	138	2	159573	urotelin II-omicron pr	5.47e+02
170	6	0.5	139	2	159574	urotelin II-pi pr	5.47e+02
171	6	0.5	140	2	159575	urotelin II-rho pr	5.47e+02
172	6	0.5	141	2	159576	urotelin II-sigma pr	5.47e+02
173	6	0.5	142	2	159577	urotelin II-tau pr	5.47e+02
174	6	0.5	143	2	159578	urotelin II-phi pr	5.47e+02
175	6	0.5	144	2	159579	urotelin II-psi pr	5.47e+02
176	6	0.5	145	2	159580	urotelin II-omega pr	5.47e+02
177	6	0.5	146	2	159581	urotelin II-alpha pr	5.47e+02
178	6	0.5	147	2	159582	urotelin II-beta pr	5.47e+02
179	6	0.5	148	2	159583	urotelin II-gamma pr	5.47e+02
180	6	0.5	149	2	159584	urotelin II-delta pr	5.47e+02
181	6	0.5	150	2	159585	urotelin II-epsilon pr	5.47e+02
182	6	0.5	151	2	159586	urotelin II-zeta pr	5.47e+02
183	6	0.5	152	2	159587	urotelin II-eta pr	5.47e+02
184	6	0.5	153	2	159588	urotelin II-theta pr	5.47e+02
185	6	0.5	154	2	159589	urotelin II-iota pr	5.47e+02
186	6	0.5	155	2	159590	urotelin II-kappa pr	5.47e+02
187	6	0.5	156	2	159591	urotelin II-lambda pr	5.47e+02
188	6	0.5	157	2	159592	urotelin II-mu pr	5.47e+02
189	6	0.5	158	2	159593	urotelin II-nu pr	5.47e+02
190	6	0.5	159	2	159594	urotelin II-xi pr	5.47e+02
191	6	0.5	160	2	159595	urotelin II-omicron pr	5.47e+02
192	6	0.5	161	2	159596	urotelin II-pi pr	5.47e+02
193	6	0.5	162	2	159597	urotelin II-rho pr	5.47e+02
194	6	0.5	163	2	159598	urotelin II-sigma pr	5.47e+02
195	6	0.5	164	2	159599	urotelin II-tau pr	5.47e+02
196	6	0.5	165	2	159600	urotelin II-phi pr	5.47e+02
197	6	0.5	166	2	159601	urotelin II-psi pr	5.47e+02
198	6	0.5	167	2	159602	urotelin II-omega pr	5.47e+02
199	6	0.5	168	2	159603	urotelin II-alpha pr	5.47e+02
200	6	0.5	169	2	159604	urotelin II-beta pr	5.47e+02
201	6	0.5	170	2	159605	urotelin II-gamma pr	5.47e+02
202	6	0.5	171	2	159606	urotelin II-delta pr	5.47e+02
203	6	0.5	172	2	159607	urotelin II-epsilon pr	5.47e+02
204	6	0.5	173	2	159608	urotelin II-zeta pr	5.47e+02
205	6	0.5	174	2	159609	urotelin II-eta pr	5.47e+02

257	6	0.5	228	2	559377	ABC transporter, ATP-	5.47e+02
258	6	0.5	229	2	559378	ABC transporter, ATP-	5.47e+02
259	6	0.5	230	2	559379	ABC transporter, ATP-	5.47e+02
260	6	0.5	231	2	559380	ABC transporter, ATP-	5.47e+02
261	6	0.5	232	2	559381	ABC transporter, ATP-	5.47e+02
262	6	0.5	233	2	559382	ABC transporter, ATP-	5.47e+02
263	6	0.5	234	2	559383	ABC transporter, ATP-	5.47e+02
264	6	0.5	235	2	559384	ABC transporter, ATP-	5.47e+02
265	6	0.5	236	2	559385	ABC transporter, ATP-	5.47e+02
266	6	0.5	237	2	559386	ABC transporter, ATP-	5.47e+02
267	6	0.5	238	2	559387	ABC transporter, ATP-	5.47e+02
268	6	0.5	239	2	559388	ABC transporter, ATP-	5.47e+02
269	6	0.5	240	2	559389	ABC transporter, ATP-	5.47e+02
270	6	0.5	241	2	559390	ABC transporter, ATP-	5.47e+02
271	6	0.5	242	2	559391	ABC transporter, ATP-	5.47e+02
272	6	0.5	243	2	559392	ABC transporter, ATP-	5.47e+02
273	6	0.5	244	2	559393	ABC transporter, ATP-	5.47e+02
274	6	0.5	245	2	559394	ABC transporter, ATP-	5.47e+02
275	6	0.5	246	2	559395	ABC transporter, ATP-	5.47e+02
276	6	0.5	247	2	559396	ABC transporter, ATP-	5.47e+02
277	6	0.5	248	2	559397	ABC transporter, ATP-	5.47e+02
278	6	0.5	249	2	559398	ABC transporter, ATP-	5.47e+02
279	6	0.5	250	2	559399	ABC transporter, ATP-	5.47e+02
280	6	0.5	251	2	559400	ABC transporter, ATP-	5.47e+02
281	6	0.5	252	2	559401	ABC transporter, ATP-	5.47e+02
282	6	0.5	253	2	559402	ABC transporter, ATP-	5.47e+02
283	6	0.5	254	2	559403	ABC transporter, ATP-	5.47e+02
284	6	0.5	255	2	559404	ABC transporter, ATP-	5.47e+02
285	6	0.5	256	2	559405	ABC transporter, ATP-	5.47e+02
286	6	0.5	257	2	559406	ABC transporter, ATP-	5.47e+02
287	6	0.5	258	2	559407	ABC transporter, ATP-	5.47e+02
288	6	0.5	259	2	559408	ABC transporter, ATP-	5.47e+02
289	6	0.5	260	2	559409	ABC transporter, ATP-	5.47e+02
290	6	0.5	261	2	559410	ABC transporter, ATP-	5.47e+02
291	6	0.5	262	2	559411	ABC transporter, ATP-	5.47e+02
292	6	0.5	263	2	559412	ABC transporter, ATP-	5.47e+02
293	6	0.5	264	2	559413	ABC transporter, ATP-	5.47e+02
294	6	0.5	265	2	559414	ABC transporter, ATP-	5.47e+02
295	6	0.5	266	2	559415	ABC transporter, ATP-	5.47e+02
296	6	0.5	267	2	559416	ABC transporter, ATP-	5.47e+02
297	6	0.5	268	2	559417	ABC transporter, ATP-	5.47e+02
298	6	0.5	269	2	559418	ABC transporter, ATP-	5.47e+02
299	6	0.5	270	2	559419	ABC transporter, ATP-	5.47e+02
300	6	0.5	271	2	559420	ABC transporter, ATP-	5.47e+02
301	6	0.5	272	2	559421	ABC transporter, ATP-	5.47e+02
302	6	0.5	273	2	559422	ABC transporter, ATP-	5.47e+02
303	6	0.5	274	2	559423	ABC transporter, ATP-	5.47e+02
304	6	0.5	275	2	559424	ABC transporter, ATP-	5.47e+02
305	6	0.5	276	2	559425	ABC transporter, ATP-	5.47e+02
306	6	0.5	277	2	559426	ABC transporter, ATP-	5.47e+02
307	6	0.5	278	2	559427	ABC transporter, ATP-	5.47e+02

206	6	0.5	178	2	151057	glucagon II precursor	5.47e+02
207	6	0.5	179	2	151058	glucagon II precursor	5.47e+02
208	6	0.5	180	2	151059	glucagon II precursor	5.47e+02
209	6	0.5	181	2	151060	glucagon II precursor	5.47e+02
210	6	0.5	182	2	151061	glucagon II precursor	5.47e+02
211	6	0.5	183	2	151062	glucagon II precursor	5.47e+02
212	6	0.5	184	2	151063	glucagon II precursor	5.47e+02
213	6	0.5	185	2	151064	glucagon II precursor	5.47e+02
214	6	0.5	186	2	151065	glucagon II precursor	5.47e+02
215	6	0.5	187	2	151066	glucagon II precursor	5.47e+02
216	6	0.5	188	2	151067	glucagon II precursor	5.47e+02
217	6	0.5	189	2	151068	glucagon II precursor	5.47e+02
218	6	0.5	190	2	151069	glucagon II precursor	5.47e+02
219	6	0.5	191	2	151070	glucagon II precursor	5.47e+02
220	6	0.5	192	2	151071	glucagon II precursor	5.47e+02
221	6	0.5	193	2	151072	glucagon II precursor	5.47e+02
222	6	0.5	194	2	151073	glucagon II precursor	5.47e+02
223	6	0.5	195	2	151074	glucagon II precursor	5.47e+02
224	6	0.5	196	2	151075	glucagon II precursor	5.47e+02
225	6	0.5	197	2	151076	glucagon II precursor	5.47e+02
226	6	0.5	198	2	151077	glucagon II precursor	5.47e+02
227	6	0.5	199	2	151078	glucagon II precursor	5.47e+02
228	6	0.5	200	2	151079	glucagon II precursor	5.47e+02
229	6	0.5	201	2	151080	glucagon II precursor	5.47e+02
230	6	0.5	202	2	151081	glucagon II precursor	5.47e+02
231	6	0.5	203	2	151082	glucagon II precursor	5.47e+02
232	6	0.5	204	2	151083	glucagon II precursor	5.47e+02
233	6	0.5	205	2	151084	glucagon II precursor	5.47e+02
234	6	0.5	206	2	151085	glucagon II precursor	5.47e+02
235	6	0.5	207	2	151086	glucagon II precursor	5.47e+02
236	6	0.5	208	2	151087	glucagon II precursor	5.47e+02
237	6	0.5	209	2	151088	glucagon II precursor	5.47e+02
238	6	0.5	210	2	151089	glucagon II precursor	5.47e+02
239	6	0.5	211	2	151090	glucagon II precursor	5.47e+02
240	6	0.5	212	2	151091	glucagon II precursor	5.47e+02
241	6	0.5	213	2	151092	glucagon II precursor	5.47e+02
242	6	0.5	214	2	151093	glucagon II precursor	5.47e+02
243	6	0.5	215	2	151094	glucagon II precursor	5.47e+02
244	6	0.5	216	2	151095	glucagon II precursor	5.47e+02
245	6	0.5	217	2	151096	glucagon II precursor	5.47e+02
246	6	0.5	218	2	151097	glucagon II precursor	5.47e+02
247	6	0.5	219	2	151098	glucagon II precursor	5.47e+02
248	6	0.5	220	2	151099	glucagon II precursor	5.47e+02
249	6	0.5	221	2	151100	glucagon II precursor	5.47e+02
250	6	0.5	222	2	151101	glucagon II precursor	5.47e+02
251	6	0.5	223	2	151102	glucagon II precursor	5.47e+02
252	6	0.5	224	2	151103	glucagon II precursor	5.47e+02
253	6	0.5	225	2	151104	glucagon II precursor	5.47e+02
254	6	0.5	226	2	151105	glucagon II precursor	5.47e+02
255	6	0.5	227	2	151106	glucagon II precursor	5.47e+02
256	6	0.5	228	2	151107	glucagon II precursor	5.47e+02

308	6	0.5	282	2	559462	pre-mRNA splicing fac	5.47e+02
309	6	0.5	283	2	559463	pre-mRNA splicing fac	5.47e+02
310	6	0.5	284				







```

971 6 0.5 2142 2 B35088 HMC class III histone 5,47e+02
972 6 0.5 2119 1 GNVY84 gene 5,47e+02
973 6 0.5 2119 1 GNVY84 gene 5,47e+02
974 6 0.5 2225 1 A23443 pyridine synthase 5,47e+02
975 6 0.5 2227 1 GNVY84 genome polypeptide 5,47e+02
976 6 0.5 2227 1 GNVY84 genome polypeptide 5,47e+02
977 6 0.5 2227 1 GNVY84 genome polypeptide 5,47e+02
978 6 0.5 2227 1 GNVY84 genome polypeptide 5,47e+02
979 6 0.5 2233 1 A39086 agrin precursor 5,47e+02
980 6 0.5 2316 2 A39086 agrin precursor 5,47e+02
981 6 0.5 2340 2 A39086 agrin precursor 5,47e+02
982 6 0.5 2428 2 A39086 agrin precursor 5,47e+02
983 6 0.5 2428 2 A39086 agrin precursor 5,47e+02
984 6 0.5 2428 2 A39086 agrin precursor 5,47e+02
985 6 0.5 2440 2 B35162 CMB-binding protein 5,47e+02
986 6 0.5 2440 2 B35162 CMB-binding protein 5,47e+02
987 6 0.5 2440 2 B35162 CMB-binding protein 5,47e+02
988 6 0.5 2514 1 GNVY84 genome polypeptide 5,47e+02
989 6 0.5 2567 2 A49551 thyroglobulin precursor 5,47e+02
990 6 0.5 2567 2 A49551 thyroglobulin precursor 5,47e+02
991 6 0.5 3421 1 W2826 357K transcript 5,47e+02
992 6 0.5 3421 1 W2826 357K transcript 5,47e+02
993 6 0.5 3432 1 GNVY84 genome polypeptide 5,47e+02
994 6 0.5 3432 1 GNVY84 genome polypeptide 5,47e+02
995 6 0.5 3432 1 GNVY84 genome polypeptide 5,47e+02
996 6 0.5 3432 1 GNVY84 genome polypeptide 5,47e+02
997 6 0.5 3898 1 GNVY84 genome polypeptide 5,47e+02
998 6 0.5 4377 2 A55575 ankyrin 3, long splice 5,47e+02
999 6 0.5 5116 2 A40450 tyrosine receptor/Ca 5,47e+02
1000 6 0.5 26926 1 A39450 titin, cardiac muscle 5,47e+02

```

## ALIGNMENTS

```

RESULT 1
ENTRY 157644 #type fragment
TITLE cDNA fragment
FORMAL_NAME Mus musculus (common name mouse)
DATE 02-Aug-1996 sequence_revision 02-Aug-1996 text_change
157644
ACCESSIONS 157644
AUTHORS Sage, Y.; Tung, J.; Shen, F.W.; Boyce, E.A.
JOURNAL Biol. Cell Biol. (1998) 8:4689-4695
ACROSS-REFERENCES MIM:609586
ACCESSION 157644
STATUS preliminary; translated from CB/MDL/DBJ
#molecule-type DNA
#residues 1224 #label RES
#cross-references CB:M23354; MID:9340890; PID:9554192

```

```

RESULT 3
ENTRY A28334 #type complete
TITLE protein-tyrosine-phosphatase (EC 3.1.3.48) Ly-5 precursor
ALTERNATE_NAMES 200K leukocyte common antigen; CD45; PTPN22; T-cell surface
CONTAINS 19-May-1989 sequence_revision 19-May-1989 text_change
DATE 19-May-1989
ACCESSIONS A28334; A28381; A60931; A35321; A28075; I54450
AUTHORS Thomas, M.L.; Reynolds, P.J.; Chao, A.; Ben-Neriah, Y.
JOURNAL Proc. Natl. Acad. Sci. U.S.A. (1987) 84:5160-5163
TITLE B-cell variant of mouse T200 (Ly-5): evidence for alternative
#molecule-type DNA
#residues 1224 #label RES
#cross-references CB:M23354; MID:9340890; PID:9554192
ACCESSION A28334
STATUS preliminary; translated from CB/MDL/DBJ
#molecule-type DNA
#residues 1224 #label RES
#cross-references CB:M23354; MID:9340890; PID:9554192

```

```

GENETICS
CLASSIFICATION Ly-5
KEYWORDS 200K leukocyte common antigen; leukocyte common
antigen-tyrosine-phosphatase homology
SYNOPSIS
Query Match 0.8; Score 9; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 1,20e-03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 10 LAPFALLD 18
QY 119 LAPFALLD 127

```

```

RESULT 2
ENTRY A28335 #type fragment
TITLE cDNA fragment
FORMAL_NAME Mus musculus (common name mouse)
DATE 19-May-1989 sequence_revision 19-May-1989 text_change
157644
ACCESSIONS 157644
AUTHORS Sage, Y.; Tung, J.S.; Shen, F.W.; Boyce, E.A.
JOURNAL Proc. Natl. Acad. Sci. U.S.A. (1987) 84:5164-5168
TITLE Alternative use of 5' exon in the specification of Ly-5
#molecule-type DNA
#residues 1224 #label RES
#cross-references CB:M23354; MID:9340890; PID:9554192
ACCESSION A28335
STATUS preliminary; translated from CB/MDL/DBJ
#molecule-type DNA
#residues 1224 #label RES
#cross-references CB:M23354; MID:9340890; PID:9554192

```







```

ENTRY      15
TITLE      hypercholesterolemia protein (crn-hpals intergenic region) -
          mouse
ORIGINISM
DATE
ACCESSIONS
REFERENCE
    #adipose
        Shimada, H.; Wakiyama, R.F.; Hiratsuka, J.; Maeda, Y.; Hirata
            K.; Sugitani, M.
            submitted to JBLD, December 1989
accession   J00282
#molecule_type DNA
#residue     1-39 ##label SHI
##sequence_1 experimental_source cr. Minomura
GENETICS
    #genome
        chloroplast
REMARKS
    #length 39 molecular_weight 4313 molecular_weight 1192
    #SD
DB
    5 SWISSSTL 11
QY 366 SPFLSTL 372
RESULT 16
ENTRY      A49196
TITLE      thyrotropin receptor - guinea pig (fragment)
ORIGINISM
DATE
ACCESSIONS
REFERENCE
    #thyroid
        Robbins, L.; Robbins, L.S.; Cone, R.D.
            Endocrinology 1992 130:1857-1861
journal
    #title
        In most brown and white adipose tissues the guinea pig
            thyrotropin receptor messenger ribonucleic acid is expressed
            in both brown and white adipose tissues in the guinea pig.
        Science-reference 1992 251:1078
accession   A49196
#status
    #molecule_type nucleic acid
    #organism guinea pig
    #experimental_source adipose tissue
    #note
        sequence extracted from MCBI backbone (MCBIP:89464)
        superficially glycoprotein repeat homology
CLASSIFICATION
    alpha-2-glycoprotein repeat homology

```



```

2-174      #product endoexyribonuclease rucv %status predicted
SOMMARY      $length 174 $molecular-weight 18356 $ecnumber 8754
              $label MVT
Query Match      0.6; Score 7; DB 2; Length 174;
Best Local Similarity 100.0%; Pred. No. 1,32e+01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Db      164 GARRGG 170
      |||||||
QY      242 GARRGG 248

RESULT      27
ENTRY      S14240
TITLE      #type complete
SOURCE      25-Sep-1994
ORIGINISM    25-Sep-1994 sequence_version 10-Nov-1995 %test_change
DATE      09-Sep-1997
ACCESSIONS    S14240
REVISIONS     S14240
AUTHORS      McElreath, T.; Wattian, P.; Brasseur, R.; Ruysschaert, J.M.;
              Cornille, G.
              Insect. Immun. (1990) 58:2840-2849
              $craos-reference K019035104
              $accession S14240
              ##status preliminary
              ##molecular_type DNA
              ##craos-reference EMBL:X57153; NID:g46601; PID:g46604
GENETICS      yop10
              $length 182 $molecular-weight 20028 $ecnumber 1492
              $gene
SOMMARY      Query Match      0.6; Score 7; DB 2; Length 182;
              Best Local Similarity 100.0%; Pred. No. 1,32e+01;
              Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Db      62 GISTSL 66
      |||||||
QY      836 GISTSL 862

RESULT      28
ENTRY      MOR3IV
TITLE      #type complete
SOURCE      myosin alkali light chain 3, ventricular and slow skeletal
ORIGINISM    NCBI: NCIV; myosin II catalytic light chain, cardiac
              muscle; ventricular myosin light chain 1
              $craos-reference X60000; NCBI: NCIV; myosin II catalytic light chain 1
              $formal_name Rattus norvegicus (common_name Norway rat

```

```

Db      105 111
Q1      735 741
Q2      735 741
Q3      735 741
Q4      735 741
Q5      735 741
Q6      735 741
Q7      735 741
Q8      735 741
Q9      735 741
Q10     735 741
Q11     735 741
Q12     735 741
Q13     735 741
Q14     735 741
Q15     735 741
Q16     735 741
Q17     735 741
Q18     735 741
Q19     735 741
Q20     735 741
Q21     735 741
Q22     735 741
Q23     735 741
Q24     735 741
Q25     735 741
Q26     735 741
Q27     735 741
Q28     735 741
Q29     735 741
Q30     735 741
Q31     735 741
Q32     735 741
Q33     735 741
Q34     735 741
Q35     735 741
Q36     735 741
Q37     735 741
Q38     735 741
Q39     735 741
Q40     735 741
Q41     735 741
Q42     735 741
Q43     735 741
Q44     735 741
Q45     735 741
Q46     735 741
Q47     735 741
Q48     735 741
Q49     735 741
Q50     735 741
Q51     735 741
Q52     735 741
Q53     735 741
Q54     735 741
Q55     735 741
Q56     735 741
Q57     735 741
Q58     735 741
Q59     735 741
Q60     735 741
Q61     735 741
Q62     735 741
Q63     735 741
Q64     735 741
Q65     735 741
Q66     735 741
Q67     735 741
Q68     735 741
Q69     735 741
Q70     735 741
Q71     735 741
Q72     735 741
Q73     735 741
Q74     735 741
Q75     735 741
Q76     735 741
Q77     735 741
Q78     735 741
Q79     735 741
Q80     735 741
Q81     735 741
Q82     735 741
Q83     735 741
Q84     735 741
Q85     735 741
Q86     735 741
Q87     735 741
Q88     735 741
Q89     735 741
Q90     735 741
Q91     735 741
Q92     735 741
Q93     735 741
Q94     735 741
Q95     735 741
Q96     735 741
Q97     735 741
Q98     735 741
Q99     735 741

```

ORGANISM cerevisiae) Saccharomyces cerevisiae  
 DATE 22-Jun-1995 sequence\_revision 06-Sep-1996 seqt\_change  
 21-Nov-1997  
 ACCESSIONS 869588  
 REFERENCE 869588  
 #submission submitted to the EMBL Data Library, August 1995  
 #description The sequence of S. cerevisiae candida 8166, 9787, 9717, and  
 lambda 3073.  
 #accession 869588  
 #molecule\_type DNA  
 #accession\_type 1-237 #label DIE  
 #cross-references EMBL:U03057; NID:9927764; PID:9927800; MIPS:IDMS33c  
 GENETICS  
 #map\_position 48  
 #length 237 #molecular\_weight 35670 #checksum 2153  
 Query Match 0.63; Score 7; DB 2; Length 237;  
 Best Local Similarity 100.0%; Pred. No. 1.32e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 177 SIKMAN 183  
 |||||||  
 QY 1067 SIKMAN 1073

RESULT 32  
 ENTRY 869588  
 #type complete  
 #entry EMBL 24K protein - human adenovirus 7  
 #contig 1  
 #formal\_name Herpesvirus h7  
 #formal\_name human adenovirus  
 #organism  
 #note  
 host Homo sapiens (man)  
 18-Feb-1997  
 #accession 869588  
 #molecule\_type DNA  
 #accession\_type 1-261 #label DJJ  
 #cross-references EMBL:U03057; NID:9927764; PID:9927800; MIPS:IDMS33c  
 GENETICS  
 #map\_position 48  
 #length 237 #molecular\_weight 35670 #checksum 2153  
 Query Match 0.63; Score 7; DB 2; Length 237;  
 Best Local Similarity 100.0%; Pred. No. 1.32e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 177 SIKMAN 183  
 |||||||  
 QY 1067 SIKMAN 1073

CLASSIFICATION  
 #superfamily adenovirus early E1A protein  
 #alternative\_application DNA binding/early protein;  
 #keywords transcription regulation, RING finger  
 #feature

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 143 LKELVIA 148  
 |||||||  
 QY 99 LKELVIA 105

RESULT 34  
 ENTRY 854352  
 #type complete  
 #entry EMBL 24K protein - human  
 #contig 1  
 #formal\_name Homo sapiens common name man  
 #organism  
 #note  
 15-Jul-1995 sequence\_revision 01-Sep-1995 seqt\_change  
 #accessions 854352  
 #reference 854352  
 #author  
 #journal  
 #title  
 #status  
 #accession  
 #molecule\_type DNA  
 #accession\_type 1-278 #label ARI  
 #cross-references CB:D94988  
 GENETICS  
 #map\_position 48  
 #length 278 #molecular\_weight 31749 #checksum 9071  
 Query Match 0.64; Score 7; DB 2; Length 278;  
 Best Local Similarity 100.0%; Pred. No. 1.32e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 230 RANUPL 236  
 |||||||  
 QY 230 RANUPL 236

RESULT 35  
 ENTRY 801727  
 #type complete  
 #entry EMBL 24K protein - human  
 #contig 1  
 #formal\_name Homo sapiens common name man  
 #organism  
 #note  
 21-Dec-1990 sequence\_revision 06-Jun-1997 seqt\_change  
 #accessions 801727  
 #reference 852466

1-163,195-261  
 #product early E1A 24K protein status predicted label  
 #length 261 #molecular\_weight 28385 #checksum 084  
 Query Match 0.63; Score 7; DB 1; Length 261;  
 Best Local Similarity 100.0%; Pred. No. 1.32e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 178 LKELVIA 184  
 |||||||  
 QY 862 LKELVIA 868

RESULT 33  
 ENTRY 876610  
 #type complete  
 #entry EMBL 24K protein - synchocystis sp. (PCC 6803)  
 #contig 1  
 #formal\_name synchocystis sp.  
 #organism  
 #note  
 25-Apr-1997 sequence\_revision 25-Apr-1997 seqt\_change  
 #accessions 876610  
 #reference 876610  
 #author  
 #journal  
 #title  
 #status  
 #accession  
 #molecule\_type DNA  
 #accession\_type 1-262 #label KAN  
 #cross-references EMBL:U03057; NID:9927764; PID:9927800; MIPS:IDMS33c  
 GENETICS  
 #map\_position 48  
 #length 262 #molecular\_weight 29916 #checksum 8162  
 Query Match 0.64; Score 7; DB 2; Length 262;  
 Best Local Similarity 100.0%; Pred. No. 1.32e+01;

CLASSIFICATION  
 #superfamily cAMP protein; response regulator homology  
 #keywords phosphoprotein  
 #feature  
 #status  
 #accession  
 #molecule\_type DNA  
 #accession\_type 1-262 #label KAN  
 #cross-references EMBL:U03057; NID:9927764; PID:9927800; MIPS:IDMS33c  
 GENETICS  
 #map\_position 48  
 #length 262 #molecular\_weight 29916 #checksum 8162  
 Query Match 0.64; Score 7; DB 2; Length 262;  
 Best Local Similarity 100.0%; Pred. No. 1.32e+01;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 230 RANUPL 236  
 |||||||  
 QY 230 RANUPL 236

RESULT 36  
 ENTRY 869537  
 #type complete  
 #entry EMBL 24K protein - Archaeoglobus fulgidus  
 #contig 1  
 #formal\_name Archaeoglobus fulgidus  
 #organism  
 #note  
 05-Dec-1997 sequence\_revision 05-Dec-1997 seqt\_change  
 #accessions 869537  
 #reference 869537  
 #author  
 #journal  
 #title  
 #status  
 #accession  
 #molecule\_type DNA  
 #accession\_type 1-278 #label DEN  
 #cross-references EMBL:U03057; NID:9927764; PID:9927800; MIPS:IDMS33c  
 GENETICS  
 #map\_position 48  
 #length 278 #molecular\_weight 31818 #checksum 9028  
 Query Match 0.64; Score 7; DB 2; Length 278;  
 Best Local Similarity 100.0%; Pred. No. 1.32e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 230 RANUPL 236  
 |||||||  
 QY 230 RANUPL 236

RESULT 37  
 ENTRY 869537  
 #type complete  
 #entry EMBL 24K protein - Archaeoglobus fulgidus  
 #contig 1  
 #formal\_name Archaeoglobus fulgidus  
 #organism  
 #note  
 05-Dec-1997 sequence\_revision 05-Dec-1997 seqt\_change  
 #accessions 869537  
 #reference 869537  
 #author  
 #journal  
 #title  
 #status  
 #accession  
 #molecule\_type DNA  
 #accession\_type 1-278 #label DEN  
 #cross-references EMBL:U03057; NID:9927764; PID:9927800; MIPS:IDMS33c  
 GENETICS  
 #map\_position 48  
 #length 278 #molecular\_weight 31818 #checksum 9028  
 Query Match 0.64; Score 7; DB 2; Length 278;  
 Best Local Similarity 100.0%; Pred. No. 1.32e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 230 RANUPL 236  
 |||||||  
 QY 230 RANUPL 236

```

#journal      C.R.: Venter, J.C.
#title        Nature (1997) 380:164-170
#description   The complete genome of the hyperthermophilic,
#organism      sulfate-reducing archaeon Pyrococcus furiosus.
#cross-references XREF:9804913
#accession    B69337
#status       preliminary, nucleic acid sequence not shown:

#molecule_type DNA
#taxids       1-280 #12611 K1E
#cross-references DB:EMBL/781: TIGR:AF2198
#length      260 #molecular_weight 33556 #ecolength 1366

Query Match      0.6%; Score 7; Db 2; Length 280;
Best Local Similarity 100.0%; Pred No: 1:32e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Db 263 PAFILSS 268
|||||||
QY 365 PAFILSS 371

RESULT 37
#ENTRY 37
#DEFINITION 521306 type complete
#ACCESSION 521306
#ORGANISM Pyrococcus furiosus
#DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #feat_change
#ACCESSIONS 521306
#REFERENCES 1-280
#AUTHORS Davies, S.S.P.; Krishnapillai, V.V.
#DESCRIPTOR submitted to the EMBL Data Library, September 1990
#DESCRIPTION DNA sequence analysis of the replication region of the
#accession 521306 archaeon euryhaline plasmid NF-3.
#status       preliminary
#molecule_type DNA
#taxids       1-296 #12611 DVY
#cross-references DB:EMBL/54540: PIR:945411
#length      296 #molecular_weight 33661 #ecolength 3503

Query Match      0.6%; Score 7; Db 2; Length 286;
Best Local Similarity 100.0%; Pred No: 1:32e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Db 252 STIJDQ 258
|||||||
QY 786 STIJDQ 792

RESULT 38

```

Mon Dec 21 09:28:27 1998

US-08-951-733-20.rpt

Page 55

```
ENTRY          title       type complete
SYNOPSIS       gnuv protein synchocystis ap. (PCC 6803)
ORIGIN        formal name Synchocystis sp.
DATE          25-Apr-1997 sequence revision 25-Apr-1997 text change
REFERENCES     PCC 6803
AUTHORS       Kameo, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanuma, E.;
              Nakamura, T.; Miyajima, N.; Hiroseawa, M.; Sugita, M.;
              Murakami, S.; Makishiki, N.; Katoh, K.; Ohmura, S.; Shioyama,
              Y.; Takeuchi, C.; Wada, T.; Matenabe, A.; Yamada, M.;
              Yasuda, M.; Tabeta, S.
JOURNAL        DNA Res. (1996) 3:109-116
FEATURES      CDNA SEQUENCE OF THE UNICELLULAR CYANOCHLOROUS SYNCHOCYSTIS SP. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
SEQUENCE-REFERENCE   Genbank accession number not shown; translation not shown
SOURCE         Genbank DDBJ/EBI/FBI accession ID:D1562127; PID:4101855; UID:P1652196
NOTES          The nucleotide sequence was submitted to the EMBL Data Library, June 1996
GENETICS      dnaV superfamily dnaV amino-terminal homology
CLASSIFICATION CLUSTAL
FEATURES      *consid dnav amino-terminal homology labelDnV
               #length 307 molecular-weight 34698 schematic 1120
Query Match    0.6% Score 7; DB 2; Length 307;
Best local similarity 100.0%; Pval. Nm. 1.32e+01;
Matched       7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB            129 WSPARP 125
             11111111
OY           294 WSPARP 300

RESULT       39
ENTRY       s16681
SYNOPSIS    hemocytic proteas - Human
ORIGIN      Homo sapiens common name man
DATE        21-Nov-1993 sequence_revision 10-Mar-1995 text_change
TEXT        17-Oct-1997
ACCESSIONS  s16681
```

Mon Dec 21 09:28:27 1998

US-08-951-733-20.rpt

Page 50

Page 50

```

Db      89          |||||||
Query Match       0.6%: Score 7; DB 2; Length 321;
Beat Local Similarity 100.0%; Pred. No. 1.13e+01;
Matches           7; Conservative    0; Mismatches   0; Indels   0; Gaps   0;

Oy      119 LARGAL 135

RESULT  41
ENTRY   P64877      *type complete
TITLE   Peptide transport system permease protein Sgh - Escherichia
FEATURES
        f12omal_dnae Escherichia coli
DATE     12-Sep-1997; sequence, revision 17-sep-1997 text_change
ACCESSION
        M64200
        M64200
AUTHORS   Battersby, F.R.; Plunkett III, G.; Bloch, C.A.; Perma, N.T.;
            Burdick, V.; Riley, M.; Collado-Vides, J.; Glaser, J.D.;
            Klibertsch, R.W.; Oedem, M.A.; Rose, D.J.; Han, S.; Shoo
            Science (1997) 277:1453-1462
            PMID:8754851
            PubMed reference: MEDLINE(8754851)
STATUS    H64877      preliminary; nucleic acid sequence not shown;
            translation not shown
MOLECULE.type DNA
SEQUENCE
        accession GB:A000227; GB:U00096; MID:g187543; PID:g187550;
        Genbank reference GB:A000227; UNIPROT:U0096
GENETICS
        *Experimental_source strain K-12, substrain MG1655
        gene
SUBSTRAT   ssgD
            length 321 molecular_weight 36038 rcdexnum 1919
QUERY MATCH
        Query Match         0.6%: Score 7; DB 2; Length 321;
        Beat Local Similarity 100.0%; Pred. No. 1.13e+01;
        Matches             7; Conservative    0; Mismatches   0; Indels   0; Gaps   0;

DB      83 LARGAL 89
        |||||||
Oy      119 LARGAL 125

RESULT  42
ENTRY   E65219
TITLE   conserved hypothetical protein MTH99A - Methanobacterium
        thermoautotrophicum (strain Delta H)

```



[illegible]

```

ACCESSIONS      556795
REFERENCE       S56793 P.M. Allfong, C
                #restriction submitted to the Protein Sequence Database, September 1995
                #accession S56795
                #molecule-type DNA
GENETICS        #cotton-reference MDL:502381 MID:g106574; PID:g106575; MIPS:x1d023c
SYNOPSIS        #SD:PT130
                #map_location length 347 molecular-weight 38812 checksum 1050

Query Match     0.64; Score 7; Dn 2; Length 347;
Best Local Similarity 100.0%; Pred. No. 1.32e+01;
Metrics         7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db              67 ST12D1_73
CY              786 ST12D1_792

RESULT          48
START          139490 4-type complete
END            139490 4-type complete
ORIGIN         139490 4-type complete
DATE           19-Jul-1995 sequence revision 19-Jul-1995 text change
ACCESSIONS     139490 S56790
REFERENCE       139490 S56790
AUTHORS         Gen Treich, P.; Reijser, V.; Vanderleyden, J.
TITLE           Sequence analysis of the Aspergillus brasiliensis exoB gene,
                #cotton-reference MID:g1293160 exocodon 4-spiromycin
                accession 139490
                #status preliminary; translated from GB/EMBL/DBJ
                #molecule-type DNA #label RBS
                #cotton-reference MDL:225478 MID:g386520 PID:g386521
GENETICS        exoB
                #restriction submitted to the Protein Sequence Database, September 1995
                #accession S56795
                #molecule-type DNA
CLASIFICACION   #SD:PT130
KEYWORDS        exoB, 4-spiromycin, homology
FEATURES        1..315
                #domain UDPglucose 4-epimerase homology 1Alb1-DUP
                #length 348 #molecular-weight 38700 #checksum 3399
                Query Match 0.64; Score 7; Dn 2; Length 348;
                Best Local Similarity 100.0%; Pred. No. 1.32e+01;

```

```

Matches 7: Conservative 0; Mismatches 0; Indels 0; Gaps
Db      342 GSASIS 348
QY      247 GSASIS 253

RESULT   49
ENTRY    A21669 #type complete
TITLE    Interleukin-8 receptor, high affinity - rabbit
ALTERNATE_NAMES FMR_receptor
ORGANISM formal_name Oryzotylagus cuniculus common_name domestic
                                                    22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change
DATE     31-Oct-1997
ACCESSIONS A21669
REFERENCES 1. Kasperk, J.S.; Thompson, R.L.; Thompson, K.M.; Pynn, H.Y.; Navarro, J.
            J. Biol. Chem. (1990) 265:20061-20064
            Molecular cloning of the FMR-Lau-Pne receptor from
            neutrophils
            accession A21669
            #molecule_type RNA
            #release_date 1995
            #release_ref 135 #label T80
            G protein-coupled receptor, glycoprotein, membrane protein;
            Neutrophil
SUMMARY   #length 354 #molecular_weight 40320 #ec_number 8472
Query Match          0 64; Score 7; DB 2; Length 354;
Beat Local Similarity 100.0%; Pfnd. No. 1,32e+01;
Matches 7: Conservative 0; Mismatches 0; Indels 0; Gaps 0
Db      335 LITRHAUT 341
QY      1104 LITRHAUT 1110

RESULT   50
ENTRY    J01231 #type complete
TITLE    Interleukin-8 receptor - rabbit
ORGANISM formal_name Oryzotylagus cuniculus common_name domestic
                                                    31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
DATE     20-Jan-1998
ACCESSIONS J01231; A46483
REFERENCES Buchmann, M.P.; Knuefer, W.E.; Kostlosky, C.; Vandenberg, T.;
            Price, V.; Lyman, S.; Gerard, N.P.; Gerard, C.; Cerretti,
            D.P.
```

```

#journal      Biochem. Biophys. Res. Commun. (1993) 179:74-789
#title        Molecular characterization of the interleukin-8 receptor.
#cross-references  MEDLINE 91378994
#molecule_type  DNA
#molecule_type  RNA
#statistics    1.355 #label BEC
#cross-references  DB M74240: RID:9165438: PID:9165439
REFERENCE
M4683  J. Kang, M. J. Rice, G. C. Wood, W. I.
#journal      J. Immunol. (1992) 149:1251-1264
#title        Characterization of complementary DNA clones encoding the
#journal      rabbit IL-8 receptor.
#cross-references  MEDLINE 9128145
#molecule_type  DNA
#statistics    M4683
#status        Preliminary
#molecule_type  mRNA
#statistics    1.955 #label LEM
#cross-references  MEDLINE 9165440: PID:9165441
#experimental_  sequence extracted from MCB backbone (NCBI#: 81526,
#note         NSB#: 91550)
#cross-references  MEDLINE 9165440: PID:9165441
#journal      G protein-coupled receptor; transmembrane protein
#molecule_type  Nucleic Acid
#statistics    1.585 #label 100001:
#cross-references  MEDLINE 9165440: PID:9165441
#journal      Best Local Similarity 100.0%; Pctd. No. 1.33e+01:
#molecule_type  Nucleic Acid
#statistics    0.649 #label 100001:
#cross-references  MEDLINE 9165440: PID:9165441
#journal      Matrixes 0; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
#statistics    0.649 #label 100001:
#cross-references  MEDLINE 9165440: PID:9165441
#journal      Db 336 #label 100001:
#cross-references  MEDLINE 9165440: PID:9165441
#journal      CY 1104 #label 100001:

```













DN DNA REPAIR: HYDROLASE, GLYCOSIDASE.  
 AC ACT SITE 148 148 GENERAL BASE (BY SIMILARITY).  
 DE SEQUENCE 305 AA: 34376 MW: 6003849 CMC33;  
 Query Match 0.7%: Score 8; DB 1: Length 305;  
 Best Local Similarity 100.0%; Pred. No. 4,71e-02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 231 RYALICE 238  
 |||||||  
 QY 105 RYALICE 112

RESULT 4  
 ID KMO5\_MSVTS STANDARD: PRT: 342 AA.  
 AC P10421;  
 DE 01-OCT-1989 (REL. 10, CREATED)  
 DE 01-OCT-1989 (REL. 11, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE MOS STRAIN/THROMBIN-PROTEIN KINASE TRANSFORMING PROTEIN (EC 2.7.1.-).  
 CC V-KOS.  
 CC METEOROLIZANT SARCINA VIRUS (CLORE 75155).  
 CC V-KOS: RNA ENVELOPED VIRUSES, POSITIVE STRAND, RETROVIRAL.  
 CC ONCOVIRINE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA PRILEP, STOKING C., STACEY A., OSTERING W.;  
 RA J. VIROL. 61:889-897(1987).  
 CC -1- SIMILARITY WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-  
 CC KINASES.  
 CC PIR: A16592, TAYMS.  
 CC PROTEIN KINASES.  
 DB PROSITE: P500107; PROTEIN KINASE, ATP, 1.  
 DB PROSITE: P500107; PROTEIN KINASE, ATP, 1.  
 DB TRANSFERASE: SERINE/THROMBIN-PROTEIN KINASE, ONCOGENE.  
 EN ATP-BINDING.  
 FT DOMAIN 63 338 PROTEIN KINASE.  
 FT NP\_BIND 69 77 ATP (BY SIMILARITY).  
 FT ACT\_SITE 188 198 BY SIMILARITY.  
 DE SEQUENCE 342 AA: 37970 MW: 1662430 CMC33;  
 Query Match 0.7%: Score 8; DB 1: Length 342;  
 Best Local Similarity 100.0%; Pred. No. 4,71e-02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 291 LAPSLTCA 298  
 |||||||  
 QY 372 LAPSLTCA 379

FT BINDING 334 334 CAMP.  
 FT BINDING 343 343 CAMP.  
 DE SEQUENCE 400 AA: 44962 MW: 17740099 CMC33;  
 Query Match 0.7%: Score 8; DB 1: Length 400;  
 Best Local Similarity 100.0%; Pred. No. 4,71e-02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 106 REEDTDR 113  
 |||||||  
 QY 461 REEDTDR 468

RESULT 5  
 ID KAP2\_HMAN STANDARD: PRT: 403 AA.  
 AC P13661; Q16823;  
 DE 01-JAN-1990 (REL. 13, CREATED)  
 DE 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
 DE 01-SEP-1995 (REL. 31, LAST ANNOTATION UPDATE)  
 DE CAMP-DEPENDENT PROTEIN KINASE TYPE II-ALPHA REGULATORY CHAIN.  
 CC PRANAHA OR PENA OR PENA2.  
 CC HOMO SAPIENS (HUMAN).  
 CC EUCARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA:  
 CC HUMAN: PRIMATE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA TISSEY-TESTIS, A.J.  
 RA OTEN O. WITTEBOST F., SCOTT J.D., HANSSON V., JAHNSEN T.;  
 RA PERS LEFT. 246:57-64(1989).  
 CC SEQUENCE OF 1-66 FROM N.A.  
 DB 109 REEDTDR 116  
 |||||||  
 QY 461 REEDTDR 468

RESULT 5  
 ID KAP2\_BOVIN STANDARD: PRT: 400 AA.  
 AC P00515;  
 DE 21-SEP-1986 (REL. 01, CREATED)  
 DE 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)  
 DE 01-SEP-1995 (REL. 31, LAST ANNOTATION UPDATE)  
 DE CAMP-DEPENDENT PROTEIN KINASE TYPE II-ALPHA REGULATORY CHAIN.  
 CC PRANAHA.  
 CC BOS TAURUS (BOVINE).  
 CC EUCARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA:  
 CC HUMAN: PRIMATE.  
 RN [1]  
 RP SEQUENCE.  
 RA TISSEY-TESTIS, A.J.  
 RA OTEN O. WITTEBOST F., SCOTT J.D., HANSSON V., JAHNSEN T.;  
 RA PERS LEFT. 246:57-64(1989).  
 CC SEQUENCE OF 1-66 FROM N.A.  
 DB 109 REEDTDR 116  
 |||||||  
 QY 461 REEDTDR 468

DB NEM: 1769310;  
 DB PROSITE: P500888; CAMP\_BINDING\_1, 2.  
 DB PROSITE: P500889; CAMP\_BINDING\_2, 2.  
 DB CAMP\_BINDING: PHOSPHORYLATION, DUBLICATION, MULTIGENE FAMILY;  
 EN ACETYLATION: 1, 134  
 FT DOMAIN 135 256  
 FT NP\_BIND 135 256 CAMP.  
 FT REPEAT 257 400  
 FT MOD\_RES 95 95 ACETYLATION.  
 FT BINDING 213 213 PHOSPHORYLATION (BY ACTIVATED ENZYME).  
 FT BINDING 213 213 CAMP.  
 Query Match 0.7%: Score 8; DB 1: Length 403;  
 Best Local Similarity 100.0%; Pred. No. 4,71e-02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 109 REEDTDR 116  
 |||||||  
 QY 461 REEDTDR 468







```

FT ACCT_SITE 179 179 BY SIMILARITY.
FT METAL 187 187 $INC (BY SIMILARITY).
50 SEQUENCE 216 AA: 24592 MW: 26200593 CMC32.
Query Match 0.6% Score 7; DB 1; Length 216;
Beet Local Similarity 100.0%; Pred. No. 5,9e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 208 NPAUSD 214
|||||||
QY 1142 NPAUSD 1148

RESULT 24
ID Y6GL.CAERL STANDARD: PRT: 357 AA.
AC P55336; Q19405; (REL. 34, CREATED)
AT 01-JUN-1986 (REL. 34, LAST ANNOTATION UPDATE)
DT 01-JUN-1986 (REL. 34, LAST ANNOTATION UPDATE)
DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DI 28 7 KD PROTEIN P13B6.1 IN CHROMOSOME X.
DR HYPOTHETICAL 28 7 KD PROTEIN P13B6.1 IN CHROMOSOME X.
DS P13B6.1 IN CHROMOSOME X.
OS CHROMATID TISSUE.
OC STRATOVIRUS; METAVIRUS; NEOPLASMS; NEOPLASM; EXCERPTA; RHADITIDA.
SC [1]
RP SEQUENCE FROM R.A.
RT 1387 AA: 1387 AA.
RL SUBMITTED: 1982.
RM SUBMITTED: 1982.
RL SUBMITTED (NOV-1995) TO DBL/GENBANK/DBD DATA BANKS.
CC 1. SIMILARITY: TO HUMAN PROTEIN D52.
CC DBL: 66105; G1072011; .
DR HUMAN ADENOVIRUS 257 AA: 26956 MW: 66679300 CMC32.
DR HYPOTHETICAL PROTEIN.
KM SEQUENCE 257 AA: 26956 MW: 66679300 CMC32;
Query Match 0.6% Score 7; DB 1; Length 257;
Beet Local Similarity 100.0%; Pred. No. 5,9e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 30 DAVIEQ 36
|||||||
QY 810 DAVIEQ 816

RESULT 25
ID E1A.DBD07 STANDARD: PRT: 261 AA.
AC P03256;
AT 21-JUN-1986 (REL. 01, CREATED)
DT 21-JUN-1986 (REL. 01, LAST ANNOTATION UPDATE)
DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DI EARLY E1A 28 KD PROTEIN (CONTAINS: EARLY E1A 24 KD PROTEIN).
DS HUMAN ADENOVIRUS TYPE 7.

```





US-08-951-733-20 -exp

Nov Dec 21 09:28:19 1998







**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**



















```

RP SEQUENCE FROM N.A.
NC TISSUE=ABERRANT ISLAND:
NC NC 97 MATCHES: 0; CONSERVATIVE: 0; Mismatches 0; Indels 0; Gaps 0
NC MODURA C. NANAKIVASHI O., NISHIMORI K., KITANO S.:
RL GENB 185:217-221(1997).
DC -+ INITIALLY BELONGS TO THE CYTOCHROME P450 FAMILY.
CC PROSITE: PS00067; P450.
DR PROSITE: PS00065; CYTOCHROME_P450.
FM OXIDOREDUCTASE, MONOOXYGENASE, ELECTRON TRANSPORT, MEMBRANE, HEME.
FM NC 508 AA; 5818cP CRG22.
FM NC 508 AA; 5818cP CRG22.
QY Query Match 0.7% Score 8: DB 13: Length 508;
Db 128 PEPCYLAT 135 Pred No. 1.57e-01:
OY 426 PEPCYLAT 433 Mismatches 0; Indels 0; Gaps 0

RESULT 16 PRELIMINARY: PROT. 680 AA.
AC TC053740:
DT 01-JAN-1998 (REBASED, 05, CREATED)
DT 01-JAN-1998 (REBASED, 05, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (REBASED, 05, LAST ANNOTATION UPDATE)
DE PRSA (FRAGMENT).
OS DISCRETELYN DISCORDERN (SLIME MOLD).
OS DISCRETELYN DISCORDERN (SLIME MOLD).
OC EUMYCETOPHYTES, PHOTOPHOTOPHYTES, SARCODINIUM, RHIZOGONIUM,
OC EUMYCETOPHYTES, PHOTOPHOTOPHYTES, SARCODINIUM, RHIZOGONIUM,
OC (1)
RC SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
DR SUBMITTED (AUG-1997) TO EMBL/GENBAN/DBJ DATA BANKS.
DR EMBL AF031983; G345315: -.
LC NON-TRAN.
SC SEQUENCE 680 AA; 76448 MW; 262B580 CRG22.

Query Match 0.7% Score 8: DB 5: Length 680;
Db Local Similarity 100.0%; Pred No. 1.57e-01:
Mismatches 0; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Db 166 REZILIN 173
OY 559 REZILIN 566

```

```

RESULT 11 PRELIMINARY: PRT: 869 M.A.
ID Q3444
DT 01-NOV-1996 (TREMSELT. 01, CREATED)
DT 01-NOV-1996 (TREMSELT. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMSELT. 05, LAST ANNOTATION UPDATE)
DT 01-JAN-1998 (TREMSELT. 05, LAST ANNOTATION UPDATE)
CM 18
CM 18
CM 18
CM 18
OC DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EPIARVOTIA METAZOA: ARTEROPODA: INSECTA: DIPTERA.
RN
RN SEQUENCE FROM N.A.
RA MEDLINE: 96194054.
RA
RA PARK W. J., LIU J., SHARP R. J., ADLER P. N.;
RA DEVELOPMENT 121:96106(1996).
RA PUBM: 96123508.
RA PUBM: 960401259. In.
RA PUBM: 960401259. In.
DB
DB SEQUENCE 869 M.A.: 97669 MW: 83958888 CRC32.
Query Match
Query Local Similarity 100.0%: Score 8. DB 5; Length 869;
Match 430
Match 8; Conservative 0; Mismatches 0; Gaps 0
DB 430 SEQUENCE 437
|||||||
Cy 251 SEQUENCE 258

```

[illegible]

```

RC STRAIN-04-2;
RX MEDLINE; 9518113;
RA HODSDON L., GOUTIER M.C.; LE CNER J.P., DE LOCHRESSE N., SPERLING L.;
RD 1996;
RW CHINAIE I.; 063825-35(1996);
RP SEQUENCE FROM N.A.
RC STRAIN=04-2;
RA HODSDON L., GOUTIER M.C.;
RB MOL. BIOL. CELL; 7:6584-7;
RD EMBL; 027510; G881410; -.
RM MATRIX PROTEIN.
RT NON TER
FT 23
SQ SEQUENCE 23 AA; 2706 MW; A872C0F4 CRC32;

Query Match      0.68; Score 7; DB 5; Length 33;
Seq Local Similarity 100.0%; Pct Id 47.0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 7 VSTINDL 13
CY 785 VSTINDL 791

RESULT 21
AID 065531; PRESEQUENCE; PRT; 37 AA.
DT 01-AUG-1998 (THEDBREFL_07_CREATED)
DT 01-AUG-1998 (THEDBREFL_07_LAST_SEQUENCE_UPDATE)
DT 01-AUG-1998 (THEDBREFL_07_LAST_ANNOTATION_UPDATE)
D7 01-AUG-1998 (THEDBREFL_07_LAST_ANNOTATION_UPDATE)
D8 ACCESSION NUMBER: F01281
D9 DESCRIPTION: RECEPTOR TUBULAR DOMAIN
OC ERKAFLOWN; PLASTIN; EBHROPHITIN; ANGIOSPERMAL; DICOTYLEDOREAE;
OC SOLANACEAE; SOLANACEAE.
RC SEQUENCE FROM N.A.
RB STRAIN=CV ALISA CAIG;
RA AGGELIS R., RANELLI S.A.;
RL SUBMITTED (MAR-1998) TO EMBL/CBIRANK/DDBJ DATA BANKS.
RD EMBL; 0100428; E238108; -.
RM MATRIX PROTEIN.
RT NON TER
FT 37
SQ SEQUENCE 37 AA; 4205 MW; 5C762110 CRC32;

Query Match      0.68; Score 7; DB 10; Length 37;
Seq Local Similarity 100.0%; Pct Id 1.47e+01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

DB	5	REFSEQS	11		
QY	364	REFSEQS	370		
RESULT	32	PRELIMINARY:	PRE:	87	MA.
AD	OS3151				
AD	OS3151				
DT	01-JUN-1988	(REBUILT, 06, CREATED)			
DT	01-JUN-1988	(REBUILT, 06, CREATED)			
DT	01-AUG-1988	(REBUILT, 07, LAST SEQUENCE UPDATE)			
DT	01-AUG-1988	(REBUILT, 07, LAST SEQUENCE UPDATE)			
D	OS3151	COMPLETE PLASMID SEQUENCE.			
D	SALMONELLA BERTH				
OC	PLASMID PERET.				
OC	PROKARYOTA, GRACILICUTES; SCYOBACTERIA; PNCUTIVITELY ANAEROBIC RODS;				
OC	ENTEROBACTERIACEAE.				
RP	SEQUENCE FROM N.A.				
RP	SEQUENCE FROM N.A.				
RL	HANES D. E., KASISY S. P., LAMPEL K.A.:				
RL	SUBMITTED (SEP-1987) TO EMBL/GENBANK/DBJ DATA BANKS.				
RL	GENBANK: J031795; G889636.				
RV	PLASMID				
SO	SEQUENCE 87 MA; 9613 MW; 30E256CA, CRC13:				
Query Match		0.63; Score 7; DB 2; Length 87;			
Percent Identical	100.0	Percent Identical	100.0		
Matches	7;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0
QY	183	LAMPSCA	58		
	1				
	1				
	183	LAMPSCA	189		
RESULT	33	PRELIMINARY:	PRE:	109	MA.
IT	OS3180				
AC	OS3180				
DT	01-NOV-1986	(REBUILT, 01, CREATED)			
DT	01-NOV-1986	(REBUILT, 01, CREATED)			
DT	01-NOV-1986	(REBUILT, 01, LAST SEQUENCE UPDATE)			
D	OS3180	COMPLETE PLASMID SEQUENCE.			
D2	HPVA (FRAGMENT)				
D2	HPVA (FRAGMENT)				
OC	STREPTOCOCCUS CRISTA.				
OC	PLASMID PESA8002.				
OC	PROKARYOTA, BACTERIA; FRIMICUTES; GRAM-POSITIVE COCCI;				
OC	STREPTOCOCCACEAE.				
RP	SEQUENCE FROM N.A.				
RP	SEQUENCE FROM N.A.				
RL	STRAHLING-55;				
RL	INFERCT. IMMUN. 64:214-211(1986).				
DR	EMBL: U65432; G1184931; -;				
DR	EMBL: U65432; G1184931; -;				
DN	PLASMID.				

```

US Dec 21 09:28:30 1998
US-08-951-733-20.rptc

FT  NON_TER      1
SQ  SEQUENCE      109 AA; 11728 MW; DD059548 CRC32;
      Query Match      0.6%; Score 7; DB 2; Length 109;
      Best Local Similarity 100.0%; Pred. No. 1.4e+01;
      Matches      7; Conservative      0; Mismatches      0; Indels      0; Gaps      0;
DB  63 GASTAL 69
      |||||||
OY  699 GASTAL 705

RESULT  24      PRELIMINARY:      PRT:      111 AA.
ID  004293
AC  004293;
DT  01-NOV-1996 (TRIMBLAE, 01, CRAFTED)
DT  01-NOV-1996 (TRIMBLAE, 01, LAST SEQUENCE UPDATE)
DT  01-NOV-1996 (TRIMBLAE, 01, LAST SEQUENCE UPDATE)
DE  HYPOPHYSICAL 11.3 NO PROTEIN.
DE  HYPOPHYSICAL 11.3 NO PROTEIN.
GN  ORF 2.
OS  HAEMULUS POLYORFHA (YASST).
OC  ERYTHROCYTES; PONEI; ASCOTCOOTIUM; HEMIASCOTCEPES.
PC  SEQUENCE FROM N.A.
RP  STRAIN=DL1.
RC  KNOTILIN A.1., SEREGIN S.A., TIKHONOVA L.P., ERITOV V.M.;
RL  SUBMITTED APR-1991 TO EMBL/GENBANK/DBS DATA BANKS.
RM  HYPOPHYSICAL PROTEIN.
SQ  SEQUENCE      111 AA; 12281 MW; 39D4893A CRC32;

Query Match      0.6%; Score 7; DB 3; Length 111;
      Best Local Similarity 100.0%; Pred. No. 1.4e+01;
      Matches      7; Conservative      0; Mismatches      0; Indels      0; Gaps      0;
DB  22 TRILAPT 28
      |||||||
OY  640 TRILAPT 646

RESULT  23
ID  053100      PRELIMINARY:      PRT:      113 AA.
AC  053100;
DT  01-NOV-1996 (TRIMBLAE, 01, CRAFTED)
DT  01-NOV-1996 (TRIMBLAE, 01, LAST SEQUENCE UPDATE)
DT  01-NOV-1996 (TRIMBLAE, 01, LAST SEQUENCE UPDATE)
DE  ORF77 (TRIMBLAE)
DE  HAEMULUS POLYORFHA (HYPOPHYSICAL SPERMATODES).
DE  HYPOPHYSICAL PROTEIN.
DE  PROKARYOTIC, CARICITICITES; ANTIHYPOPHYSICAL; POIBLE BACTERIA;
DE  RHODOSPIRILLACAE.
GN  (1)

```

Mon Dec 21 09:28:30 1998

US-08-951-733-20.rpt

Page 43

Mon Dec 21 09:28:30 1998

US-08-951-733-20.rap

Page 44

```

RP SEQUENCE FROM N.A.
RC STRAIN# 2.4.11.
RC MEDLINE# 9606293.
RC J. BACTERIOL. 137:6432-6431(1995).
RL J. BACTERIOL. 137:6432-6431(1995).
DR EMBL: U35443; G107069; -.
FT NAME: 1
FT NON-TER 1 113
S0 SEQUENCE 113 AA: 11864 MW: 08479AA1 CRC32:
Query Match 0.6% Score 7; DB 2; Length 113;
Best Local Similarity 100.0%; Pred No. 1476+01;
Matches 0; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Db 45 AARGSEA 51
OY 132 AARGSEA 158
RESULT 26 PRELIMINARY; PRT: 117 AA.
AC Q30000.
AC Q30002.
D7 01-NOV-1996 (TRIMBLIN. 01, CREATED)
D7 01-NOV-1996 (TRIMBLIN. 01, LAST SEQUENCE UPDATE)
D7 01-NOV-1996 (TRIMBLIN. 01, LAST SEQUENCE UPDATE)
D2 01740W-1996 (TRIMBLIN. 01, LAST ANNOTATION UPDATE)
OC MYCOPLASMA LEPTAE.
OC PNEUMONIA. FIMBRIACE; ACTINOTETIDALES; MYCOBACTERIAE.
OC 1. SEQUENCE FROM N.A.
RA SMITH D. R.;
RL SUBMITTED (APR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RA 12. SEQUENCE FROM N.A.
RL SUBMITTED (APR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RA ROBINSON F.;
RL EMBL: U11813; G593457;
DR DO SEQUENCE 117 AA: 12770 MW: BAA264E8 CRC32:
Query Match 0.6% Score 7; DB 2; Length 117;
Best Local Similarity 100.0%; Pred No. 1476+01;
Matches 0; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Db 65 AARGSEA 71
OY 243 AARGSEA 249
RESULT 27 PRELIMINARY; PRT: 115 AA.
AC Q73111.
AC Q73111.

```

[illegible]

DR EMBL D64005: C1006407: -  
 SS HUMAN T-CELL LYMPHOTROPIC VIRUS TYPE II.  
 SEQUENCE 149 AA: 14443 MW: 20040555 CRC32:  
 Query Match 0.6% Score 7: DB 2: Length 128:  
 Best Local Similarity 100.0%: Pred. No. 1.47e+01:  
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 Db 46 AVIIDS 52  
 |||||  
 QY 811 AVIIDS 817

RESULT 29  
 ID 080839 PRELIMINARY: PRT: 135 AA.  
 AC 080839: -  
 DT 01-NOV-1996 (TRIMBLAEL. 01, CREATED)  
 DT 01-NOV-1996 (TRIMBLAEL. 01, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1996 (TRIMBLAEL. 01, LAST SEQUENCE UPDATE)  
 DE T-CELL LYMPHOTROPIC VIRUS TYPE II (CLONE NO-6) TAX-2 (FRAGMENT).  
 OS HUMAN T-CELL LYMPHOTROPIC VIRUS TYPE II.  
 OC RETROVIRIDAE: ONCOVIRINAE: TYPE C ONCOVIRUS GROUP.  
 OC RETROVIRIDAE: ONCOVIRINAE: TYPE C ONCOVIRUS GROUP.  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92260186.  
 RA LUTHELE B., CHANET R.,  
 DR EMBL M63885: G139356: -  
 FT NON\_TER 1  
 SQ SEQUENCE 135 AA: 14124 MW: 18654028 CRC32:  
 Query Match 0.6% Score 7: DB 14: Length 135:  
 Best Local Similarity 100.0%: Pred. No. 1.47e+01:  
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 Db 72 SRSIPLP 78  
 |||||  
 QY 251 SRSIPLP 257

RESULT 30  
 ID 080833 PRELIMINARY: PRT: 138 AA.  
 AC 080833: -  
 DT 01-NOV-1996 (TRIMBLAEL. 01, CREATED)  
 DT 01-NOV-1996 (TRIMBLAEL. 01, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1996 (TRIMBLAEL. 01, LAST SEQUENCE UPDATE)  
 DE T-CELL LYMPHOTROPIC VIRUS TYPE II (CLONE NO-6) TAX-2 (FRAGMENT).  
 OS HUMAN T-CELL LYMPHOTROPIC VIRUS TYPE II.  
 OC RETROVIRIDAE: ONCOVIRINAE: TYPE C ONCOVIRUS GROUP.  
 OC RETROVIRIDAE: ONCOVIRINAE: TYPE C ONCOVIRUS GROUP.  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92260186.  
 RA LUTHELE B., CHANET R.,  
 DR EMBL M63885: G139356: -  
 FT NON\_TER 1  
 SQ SEQUENCE 135 AA: 14124 MW: 18654028 CRC32:  
 Query Match 0.6% Score 7: DB 14: Length 135:  
 Best Local Similarity 100.0%: Pred. No. 1.47e+01:  
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 Db 72 SRSIPLP 78  
 |||||  
 QY 251 SRSIPLP 257

RESULT 31  
 ID 082408 PRELIMINARY: PRT: 149 AA.  
 AC 082408: -  
 DT 01-NOV-1996 (TRIMBLAEL. 01, CREATED)  
 DT 01-NOV-1996 (TRIMBLAEL. 01, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1996 (TRIMBLAEL. 01, LAST SEQUENCE UPDATE)  
 DE T-CELL LYMPHOTROPIC VIRUS TYPE II (CLONE NO-6) TAX-2 (FRAGMENT).  
 OS HUMAN T-CELL LYMPHOTROPIC VIRUS TYPE II.  
 OC RETROVIRIDAE: ONCOVIRINAE: TYPE C ONCOVIRUS GROUP.  
 OC RETROVIRIDAE: ONCOVIRINAE: TYPE C ONCOVIRUS GROUP.  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 96190544.  
 RA OLIVERIA M.P., LAURENTO P., ISHAK M., ACEVEDO V., HANNESENHILAN N.,  
 DR EMBL M63885: G1463044: -  
 FT NON\_TER 1  
 SQ SEQUENCE 149 AA: 15702 MW: 4830545 CRC32:  
 Query Match 0.6% Score 7: DB 14: Length 149:  
 Best Local Similarity 100.0%: Pred. No. 1.47e+01:  
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 Db 86 SRSIPLP 92  
 |||||  
 QY 251 SRSIPLP 257

RESULT 32  
 ID 082405 PRELIMINARY: PRT: 149 AA.  
 AC 082405: -  
 DT 01-NOV-1996 (TRIMBLAEL. 01, CREATED)  
 DT 01-NOV-1996 (TRIMBLAEL. 01, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1996 (TRIMBLAEL. 01, LAST SEQUENCE UPDATE)  
 DE T-CELL LYMPHOTROPIC VIRUS TYPE II (CLONE NO-6) TAX-2 (FRAGMENT).  
 OS HUMAN T-CELL LYMPHOTROPIC VIRUS TYPE II.  
 OC RETROVIRIDAE: ONCOVIRINAE: TYPE C ONCOVIRUS GROUP.  
 OC RETROVIRIDAE: ONCOVIRINAE: TYPE C ONCOVIRUS GROUP.  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 96190544.  
 RA OLIVERIA M.P., LAURENTO P., ISHAK M., ACEVEDO V., HANNESENHILAN N.,  
 DR EMBL M63885: G1463044: -  
 FT NON\_TER 1  
 SQ SEQUENCE 149 AA: 15702 MW: 4830545 CRC32:  
 Query Match 0.6% Score 7: DB 14: Length 149:  
 Best Local Similarity 100.0%: Pred. No. 1.47e+01:  
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 Db 86 SRSIPLP 92  
 |||||  
 QY 251 SRSIPLP 257

OS HUMAN T-CELL LYMPHOTROPIC VIRUS TYPE II.  
 OC RETROVIRIDAE: ONCOVIRINAE: TYPE C ONCOVIRUS GROUP.  
 OC RETROVIRIDAE: ONCOVIRINAE: TYPE C ONCOVIRUS GROUP.  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 96190544.  
 RA OLIVERIA M.P., LAURENTO P., ISHAK M., ACEVEDO V., HANNESENHILAN N.,  
 DR EMBL M63885: G1463044: -  
 FT NON\_TER 1  
 SQ SEQUENCE 149 AA: 15702 MW: 4830545 CRC32:  
 Query Match 0.6% Score 7: DB 14: Length 149:  
 Best Local Similarity 100.0%: Pred. No. 1.47e+01:  
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 Db 86 SRSIPLP 92  
 |||||  
 QY 251 SRSIPLP 257

RESULT 31  
 ID 046381 PRELIMINARY: PRT: 146 AA.  
 AC 046381: -  
 DT 01-NOV-1996 (TRIMBLAEL. 01, CREATED)  
 DT 01-NOV-1996 (TRIMBLAEL. 01, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1996 (TRIMBLAEL. 01, LAST SEQUENCE UPDATE)  
 DE T-CELL LYMPHOTROPIC VIRUS TYPE II (CLONE NO-6) TAX-2 (FRAGMENT).  
 OS HUMAN T-CELL LYMPHOTROPIC VIRUS TYPE II.  
 OC RETROVIRIDAE: ONCOVIRINAE: TYPE C ONCOVIRUS GROUP.  
 OC RETROVIRIDAE: ONCOVIRINAE: TYPE C ONCOVIRUS GROUP.  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 96190544.  
 RA OLIVERIA M.P., LAURENTO P., ISHAK M., ACEVEDO V., HANNESENHILAN N.,  
 DR EMBL M63885: G1463044: -  
 FT NON\_TER 1  
 SQ SEQUENCE 146 AA: 16712 MW: 53345260 CRC32:  
 Query Match 0.6% Score 7: DB 2: Length 146:  
 Best Local Similarity 100.0%: Pred. No. 1.47e+01:  
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 Db 58 LABELING 64  
 |||||  
 QY 372 LABELING 378

RESULT 32  
 ID 046381 PRELIMINARY: PRT: 146 AA.  
 AC 046381: -  
 DT 01-NOV-1996 (TRIMBLAEL. 01, CREATED)  
 DT 01-NOV-1996 (TRIMBLAEL. 01, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1996 (TRIMBLAEL. 01, LAST SEQUENCE UPDATE)  
 DE T-CELL LYMPHOTROPIC VIRUS TYPE II (CLONE NO-6) TAX-2 (FRAGMENT).  
 OS HUMAN T-CELL LYMPHOTROPIC VIRUS TYPE II.  
 OC RETROVIRIDAE: ONCOVIRINAE: TYPE C ONCOVIRUS GROUP.  
 OC RETROVIRIDAE: ONCOVIRINAE: TYPE C ONCOVIRUS GROUP.  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 96190544.  
 RA OLIVERIA M.P., LAURENTO P., ISHAK M., ACEVEDO V., HANNESENHILAN N.,  
 DR EMBL M63885: G1463044: -  
 FT NON\_TER 1  
 SQ SEQUENCE 146 AA: 16712 MW: 53345260 CRC32:  
 Query Match 0.6% Score 7: DB 2: Length 146:  
 Best Local Similarity 100.0%: Pred. No. 1.47e+01:  
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 Db 58 LABELING 64  
 |||||  
 QY 372 LABELING 378

RESULT 33  
 ID 082405 PRELIMINARY: PRT: 149 AA.  
 AC 082405: -  
 DT 01-NOV-1996 (TRIMBLAEL. 01, CREATED)  
 DT 01-NOV-1996 (TRIMBLAEL. 01, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1996 (TRIMBLAEL. 01, LAST SEQUENCE UPDATE)  
 DE T-CELL LYMPHOTROPIC VIRUS TYPE II (CLONE NO-6) TAX-2 (FRAGMENT).  
 OS HUMAN T-CELL LYMPHOTROPIC VIRUS TYPE II.  
 OC RETROVIRIDAE: ONCOVIRINAE: TYPE C ONCOVIRUS GROUP.  
 OC RETROVIRIDAE: ONCOVIRINAE: TYPE C ONCOVIRUS GROUP.  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 96190544.  
 RA OLIVERIA M.P., LAURENTO P., ISHAK M., ACEVEDO V., HANNESENHILAN N.,  
 DR EMBL M63885: G1463044: -  
 FT NON\_TER 1  
 SQ SEQUENCE 149 AA: 15702 MW: 4830545 CRC32:  
 Query Match 0.6% Score 7: DB 14: Length 149:  
 Best Local Similarity 100.0%: Pred. No. 1.47e+01:  
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 Db 86 SRSIPLP 92  
 |||||  
 QY 251 SRSIPLP 257

RESULT 34  
 ID 082405 PRELIMINARY: PRT: 149 AA.  
 AC 082405: -  
 DT 01-NOV-1996 (TRIMBLAEL. 01, CREATED)  
 DT 01-NOV-1996 (TRIMBLAEL. 01, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1996 (TRIMBLAEL. 01, LAST SEQUENCE UPDATE)  
 DE T-CELL LYMPHOTROPIC VIRUS TYPE II (CLONE NO-6) TAX-2 (FRAGMENT).  
 OS HUMAN T-CELL LYMPHOTROPIC VIRUS TYPE II.  
 OC RETROVIRIDAE: ONCOVIRINAE: TYPE C ONCOVIRUS GROUP.  
 OC RETROVIRIDAE: ONCOVIRINAE: TYPE C ONCOVIRUS GROUP.  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 96190544.  
 RA OLIVERIA M.P., LAURENTO P., ISHAK M., ACEVEDO V., HANNESENHILAN N.,  
 DR EMBL M63885: G1463044: -  
 FT NON\_TER 1  
 SQ SEQUENCE 149 AA: 15702 MW: 4830545 CRC32:  
 Query Match 0.6% Score 7: DB 14: Length 149:  
 Best Local Similarity 100.0%: Pred. No. 1.47e+01:  
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 Db 86 SRSIPLP 92  
 |||||  
 QY 251 SRSIPLP 257

AC 002405;  
 DT 01-NOV-1996 (TREMBL, 01, CREATED)  
 DT 01-NOV-1996 (TREMBL, 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBL, 06, LAST ANNOTATION UPDATE)  
 DE T-CELL LYMPHOTROPIC VIRUS TYPE 2 REX AND TAX GENES, PARTIAL CDS, CLONE CM REX PA (FRAGMENT).  
 CM REX PA (FRAGMENT).  
 CC HUMAN T-CELL LYMPHOTROPIC VIRUS TYPE II.  
 CC VIRIDAE: SS-RNA NONENVELOPED VIRUSES; POSITIVE STRAND RNA VIRUSES.  
 CC RETROVIRAL: ONCOVIRINAE: TYPE C ONCOVIRUS GROUP;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN: N. NOVOA P., MONTE C., OLIVERA M.D., OLIVERA O.D., ISHAK R., ZHU S.W., KIRO T., HALL W.W.  
 RA J. VIROL. 70:1481-1492(1996).  
 DR EMBL: U12875; G1463047;  
 FT NON-TER 1 15779 MW; 58356CB3 CRC32;  
 SQ SEQUENCE 149 AA: 15779 MW; 58356CB3 CRC32;  
 Query Match  
 Best Local Similarity 100.0%; Pred. No. 1,47e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 86 SRSIDUP 92  
 OY 251 SRSIDUP 257

DT 01-NOV-1996 (TREMBL, 01, CREATED)  
 DT 01-NOV-1996 (TREMBL, 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBL, 06, LAST ANNOTATION UPDATE)  
 DE HEAT-SHOCK COGNATE (FRAGMENT).  
 DE DADGUS CAROTA (CAROT).  
 CC EMBRYONIC PLANT: EMBRYONIC; ANGIOSPERMAE; DICOTYLEDONAE; APiales.  
 CC UMBELLIFERAE.  
 RP SEQUENCE FROM N.A.  
 RA STRAIN: DAVEN HALF-LONG, TISSUE-SOMATIC EMBRYOS AT THE GLOBULAR STAGE;  
 RA LIN X., HANX G.J., ZIMMERMAN J.T.;  
 RA SUBMITTED NAME: G123456 TO EMBL/GENBANK/DBJ DATA BANKS.  
 RA EMBL: U12345; G123456;  
 DR EMBL: PRO0011; HSP20;  
 FT NON-TER 1  
 SQ SEQUENCE 153 AA: 18394 MW; A7D5C5A5 CRC32;  
 Query Match  
 Best Local Similarity 100.0%; Pred. No. 1,47e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 23 LPRIDR 29  
 OY 401 LPRIDR 407

RE J. VIROL. 70:1481-1492(1996).  
 DR EMBL: U12875; G1463047;  
 FT NON-TER 1  
 SQ SEQUENCE 149 AA: 15764 MW; 7BADA4AC CRC32;  
 Query Match  
 Best Local Similarity 100.0%; Pred. No. 1,47e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 86 SRSIDUP 92  
 OY 251 SRSIDUP 257

DT 01-NOV-1996 (TREMBL, 01, CREATED)  
 DT 01-NOV-1996 (TREMBL, 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBL, 06, LAST ANNOTATION UPDATE)  
 DE STRAIN: DAVEN HALF-LONG, TISSUE-SOMATIC EMBRYOS AT THE GLOBULAR STAGE;  
 DE STRAIN: DAVEN HALF-LONG, TISSUE-SOMATIC EMBRYOS AT THE GLOBULAR STAGE;  
 CC EMBRYONIC PLANT: EMBRYONIC; ANGIOSPERMAE; DICOTYLEDONAE; APiales.  
 CC UMBELLIFERAE.  
 RP SEQUENCE FROM N.A.  
 RA STRAIN: DAVEN HALF-LONG, TISSUE-SOMATIC EMBRYOS AT THE GLOBULAR STAGE;  
 RA LIN X., HANX G.J., ZIMMERMAN J.T.;  
 RA SUBMITTED NAME: G123456 TO EMBL/GENBANK/DBJ DATA BANKS.  
 RA EMBL: PRO0011; HSP20;  
 FT NON-TER 1  
 SQ SEQUENCE 153 AA: 18394 MW; A7D5C5A5 CRC32;  
 Query Match  
 Best Local Similarity 100.0%; Pred. No. 1,47e+01;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 23 LPRIDR 29  
 OY 401 LPRIDR 407

DB EMBL Y13051: E1263101: -  
 AC 085601: PRELIMINARY: PRT: 170 AA.  
 SEQUENCE 170 AA: 18402 MW: 8054442 CRC32:  
 Query Match 0.64: Score 7: DB 14: Length 170:  
 Best Local Similarity 100.0%: Pred. No. 1.47e+01:  
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:  
 DB 107 SHELUP 113  
 AC 107 SHELUP 257  
 SEQUENCE 257 AA: 251 SHELUP 257  
 Query Match 0.64: Score 7: DB 14: Length 170:  
 Best Local Similarity 100.0%: Pred. No. 1.47e+01:  
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DB EMBL Y13051: E1263101: -  
 AC 085601: PRELIMINARY: PRT: 170 AA.  
 SEQUENCE 170 AA: 18402 MW: 8054442 CRC32:  
 Query Match 0.64: Score 7: DB 14: Length 170:  
 Best Local Similarity 100.0%: Pred. No. 1.47e+01:  
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DB 107 SHELUP 113  
 AC 107 SHELUP 257  
 SEQUENCE 257 AA: 251 SHELUP 257  
 Query Match 0.64: Score 7: DB 14: Length 170:  
 Best Local Similarity 100.0%: Pred. No. 1.47e+01:  
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

DB 107 SHELUP 113  
 AC 107 SHELUP 257  
 SEQUENCE 257 AA: 251 SHELUP 257  
 Query Match 0.64: Score 7: DB 14: Length 170:  
 Best Local Similarity 100.0%: Pred. No. 1.47e+01:  
 Matches 7: Conservative 0: Mismatches 0: Indels 0: Gaps 0:



**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**